LA-UR-19-21274

Approved for public release; distribution is unlimited.

Title: Accelerated modeling of atomistic physics with machine learning

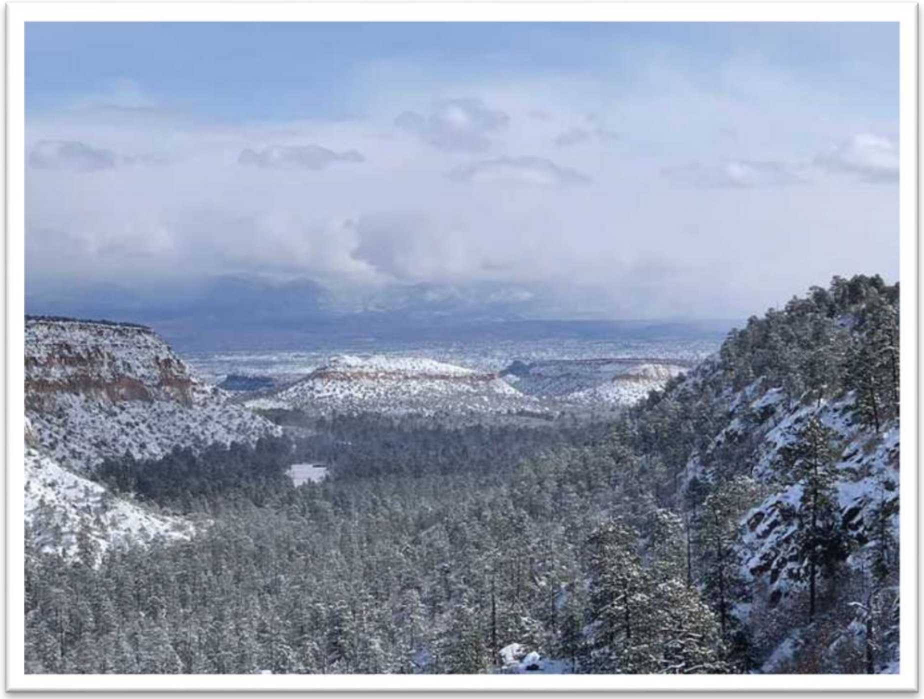
Author(s): Smith, Justin Steven; Lubbers, Nicholas Edward; Barros, Kipton Marcos; Nebgen, Benjamin Tyler; Tretiak, Sergei; Germann, Timothy Clark; Fensin, Saryu Jindal; Roitberg, Adrian E.; Isayev, Olexandr; zubatyuk, roman; Burakovsky, Leonid; Devereux, Christian; Ranashingha, Kavindri; Suwa, Hidemaro; Batista, Christian; Chern, Gai-Wei

Intended for: Machine Learning for Computational Fluid and Solid Dynamics, 2019-02-19/2019-02-21 (Santa Fe, New Mexico, United States)

Issued: 2019-02-19

Accelerated modeling of atomistic physics with machine learning

Justin S. Smith



LANL Team

Benjamin Nebgen
Kipton Barros
Saryu Fensin
Tim Germann
Leonid Burakovsky
Nicholas Lubbers
Sergei Tretiak

Collaborators

Olexandr Isayev
Adrian Roitberg
Roman Zubatyuk
Christian Devereux
Kavindri Ranashingha
Hidemaro Suwa
Christian Batista
Gia-Wei Chern

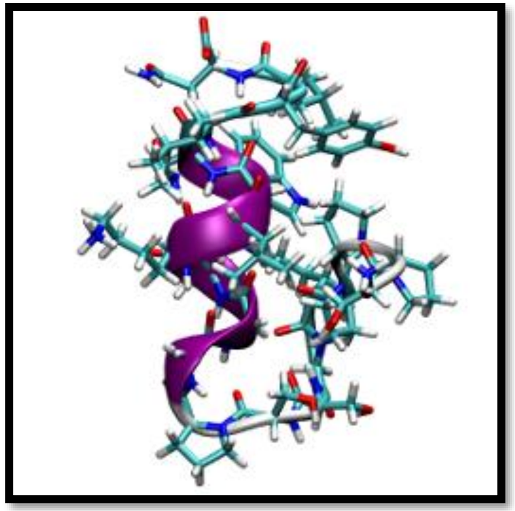
Machine Learning for Computational Fluid and Solid Dynamics

Los Alamos NATIONAL LABORATORY

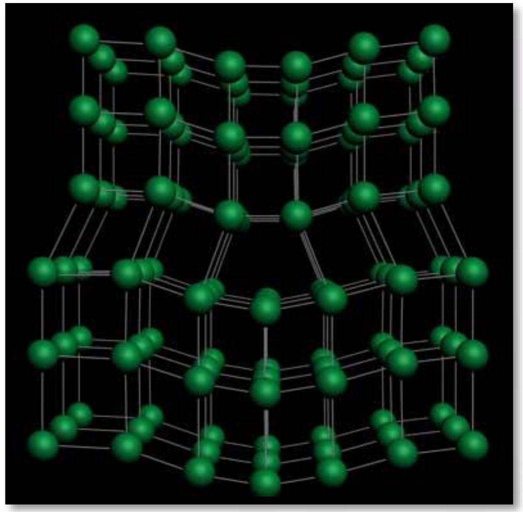
Center for Nonlinear Studies EST. 1980

Molecular (atomistic) dynamics

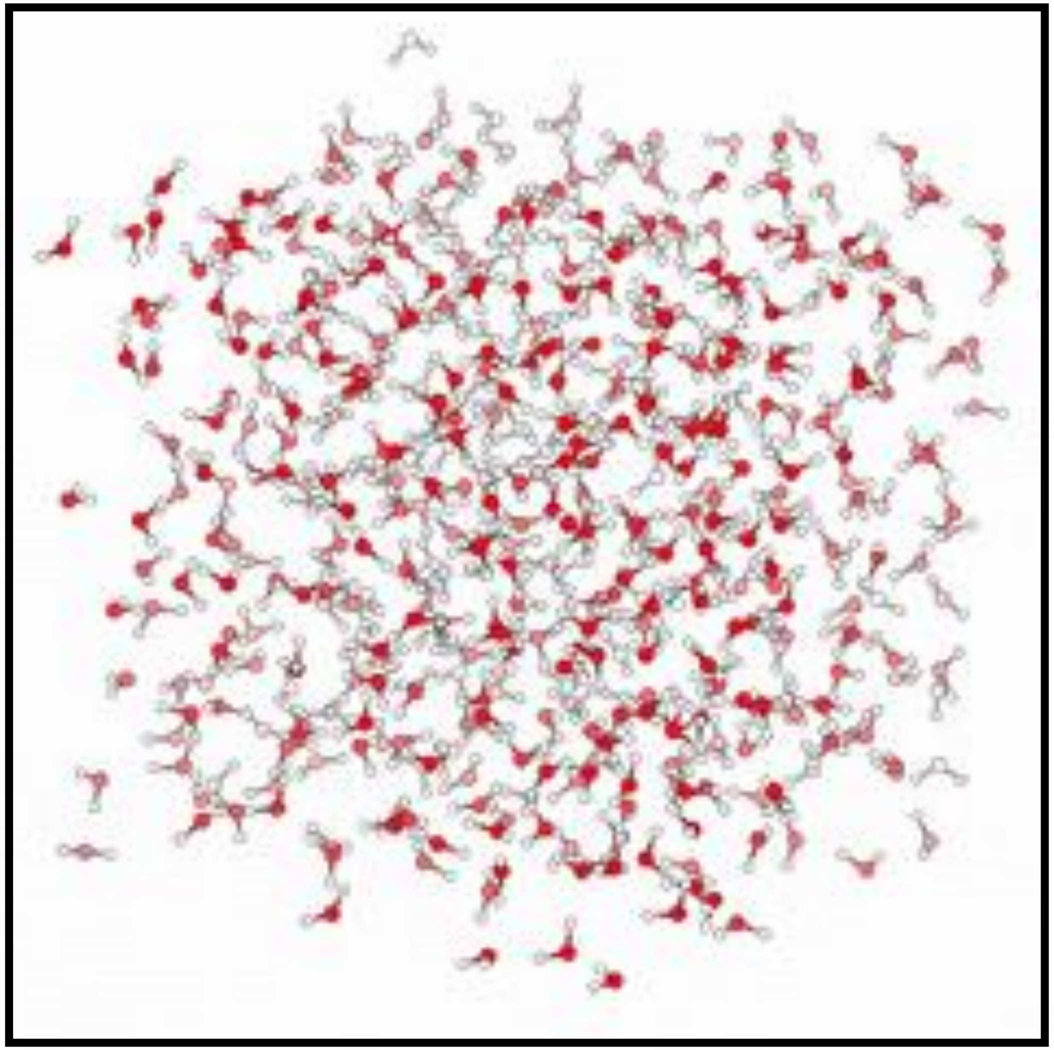
Proteins



Materials



Liquids



Energy $E[\mathbf{r}_1, \mathbf{r}_2, \ldots]$

Force $\mathbf{f}_i = -\nabla_i E$

In principle, requires a quantum mechanical calculation at *each* time step!

Dynamics $m\dfrac{d^2\mathbf{r}_i}{dt^2} = \mathbf{f}_i$

Molecular mechanics/Classical force field	The Electronic Schrödinger Equation (QM)

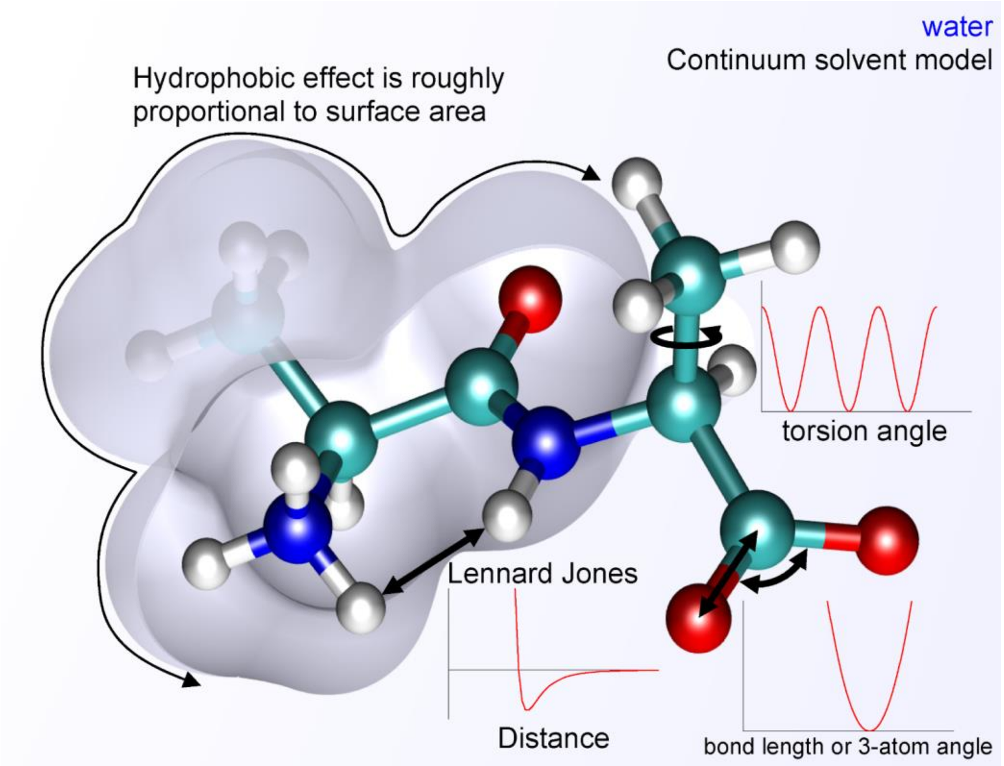
Molecular mechanics/Classical force field

Pros:

- Computationally efficient
- Accurate on systems in fitting set

Cons:

- Not very transferable
- Non-reactive
- Difficult reparameterization



https://en.wikipedia.org/wiki/Force_field_(chemistry)

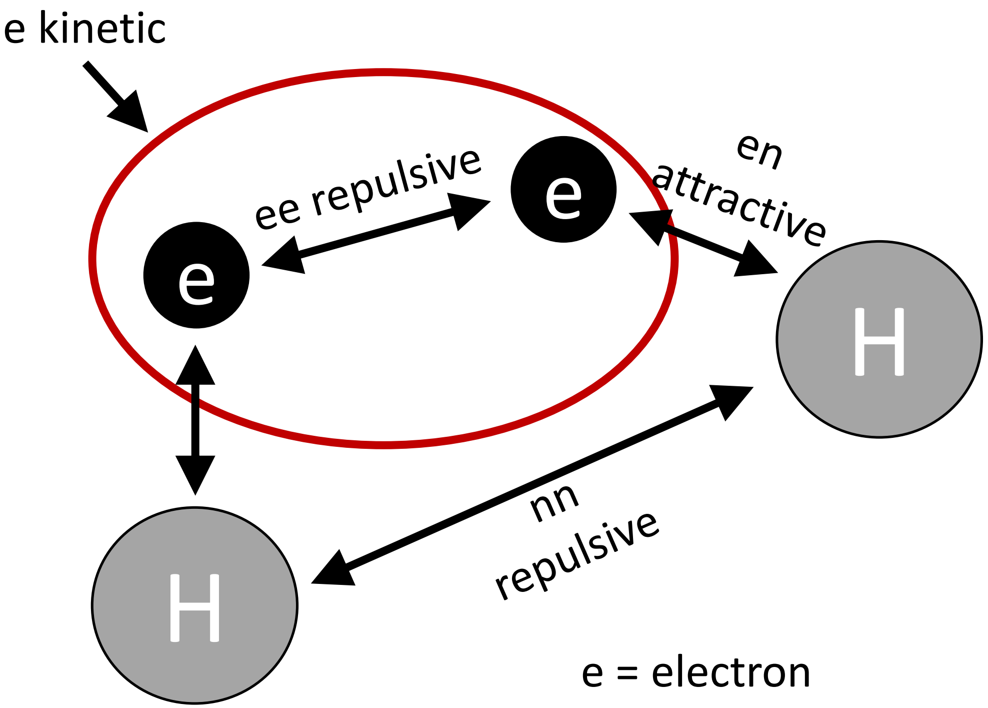
The Electronic Schrödinger Equation (QM)

Pros:

- Transferable
- Accurate

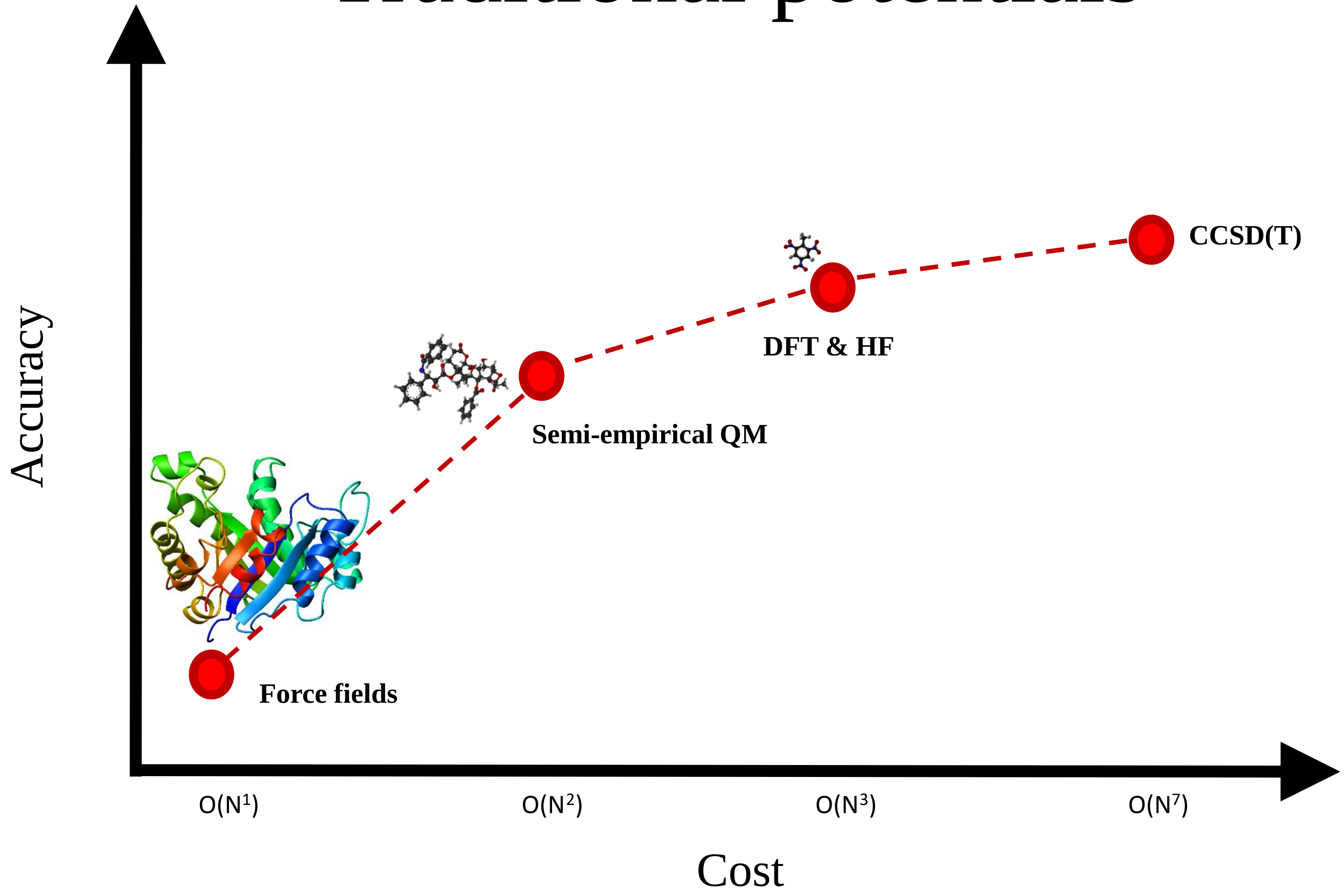
Cons:

- Computationally demanding



e = electron
n = nucleus

Traditional potentials



A potential solution

Solution: Develop an empirical potential that is accurate, fast and parametrizes itself

- *Machine learning* provides methods that fit this need

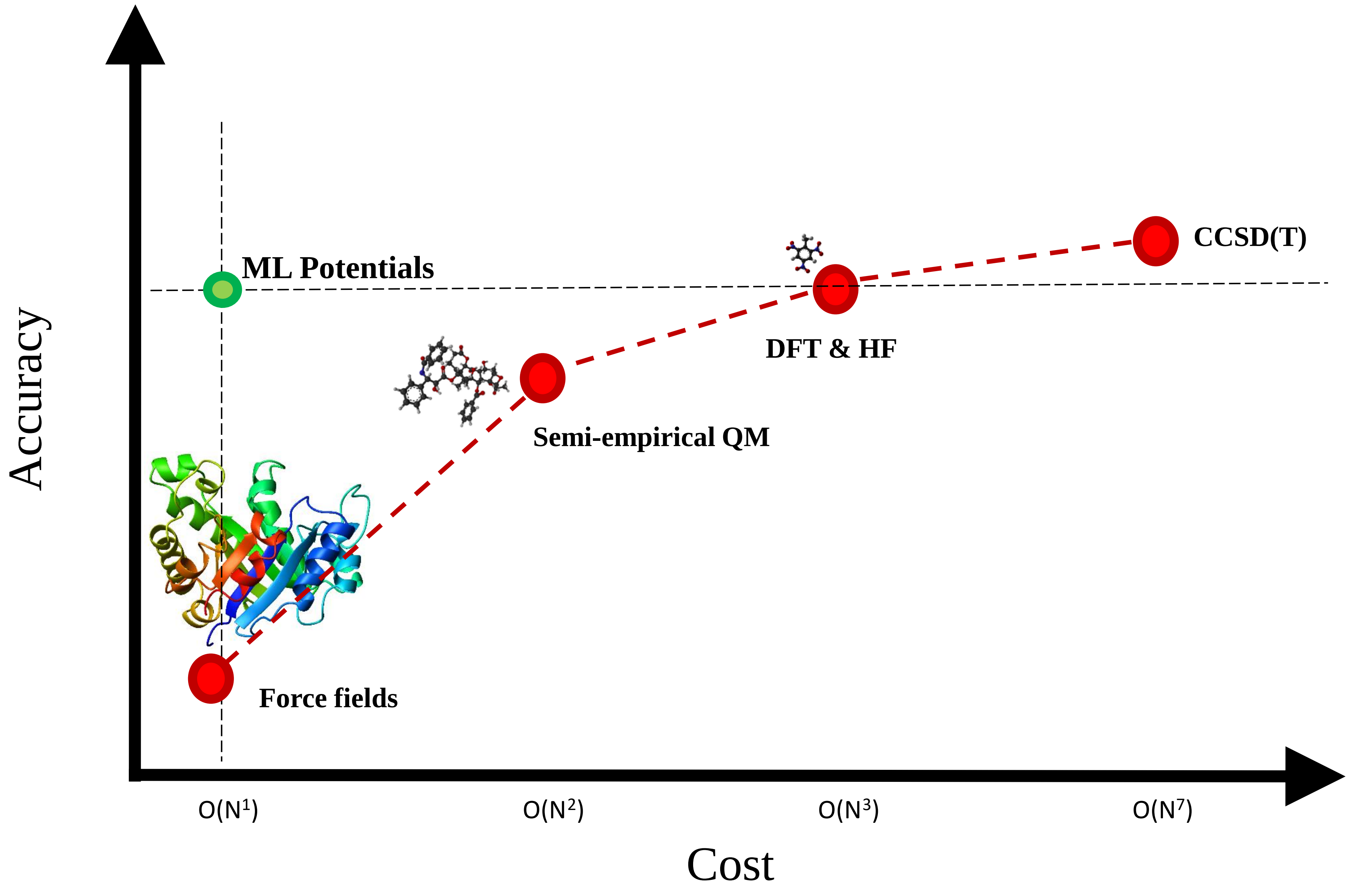
- Prior neural network potentials* (NNP) for organic molecules and materials…
  - are trained to specific molecules or phases of a material
  - are non-transferable

- Our goal: build general and accurate ML potentials

  Combine big data and deep learning concepts and a new molecular representation to produce *accurate*, *transferable*, *and extensible* NNPs.

*Behler, J.; *Angew. Chemie Int. Ed.* **2017**, *56* (42), 12828–12840.

Where does ML fit?



Design principles for ML potentials

<u>Mapping from coordinates R → Energy (& Forces) but with no a-priori functional form</u>

- Fast, accurate, and reproducible
- Reactive
- No “atom typing” required
- Conserves energy (in MD)
- Extensible to new, larger systems of atoms
- Highly automated parametrization
- Systematically improvable

Coordinates and atomic numbers $(\vec{r}, Z)$

↓

Mapping Function $f(\vec{r}, Z)$

↓

Energies E

↓

Forces $\vec{F} = -\vec{\nabla} E$

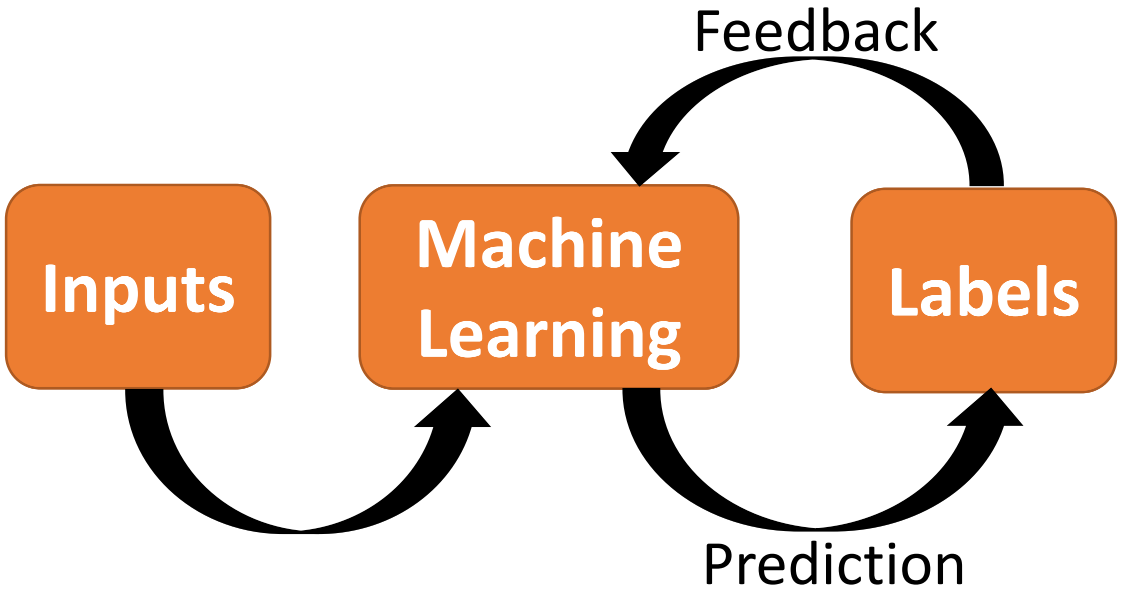
Machine learning basics

Training dataset for supervised learning

Inputs = $\{x_0, x_1, x_2, \dots, x_N\}$

Labels = $\{y_0, y_1, y_2, \dots, y_N\}$

Supervised Learning



Types of tasks

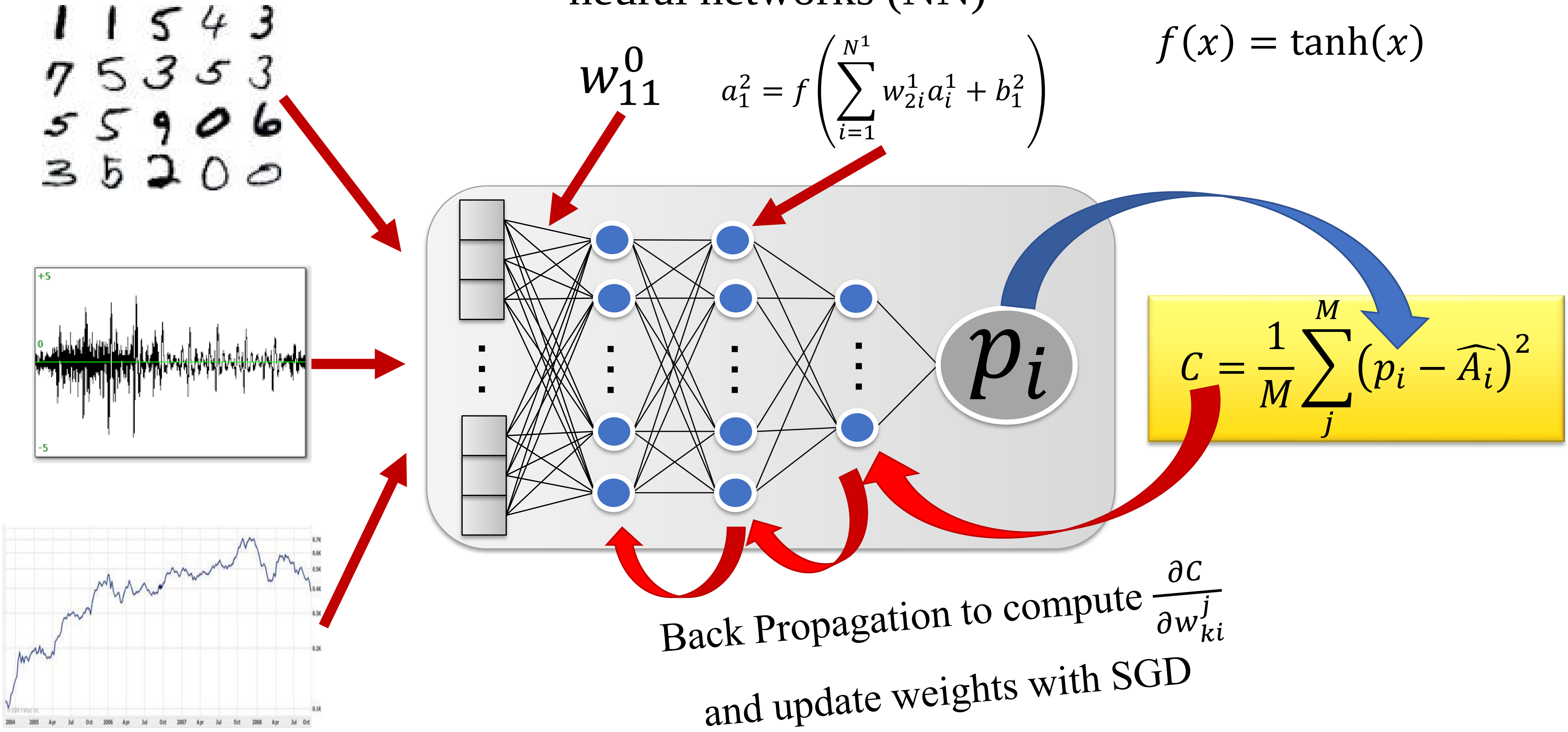
- Regression
- Classification

A few applications

- Image recognition
- Social media moderation
- Stock market prediction

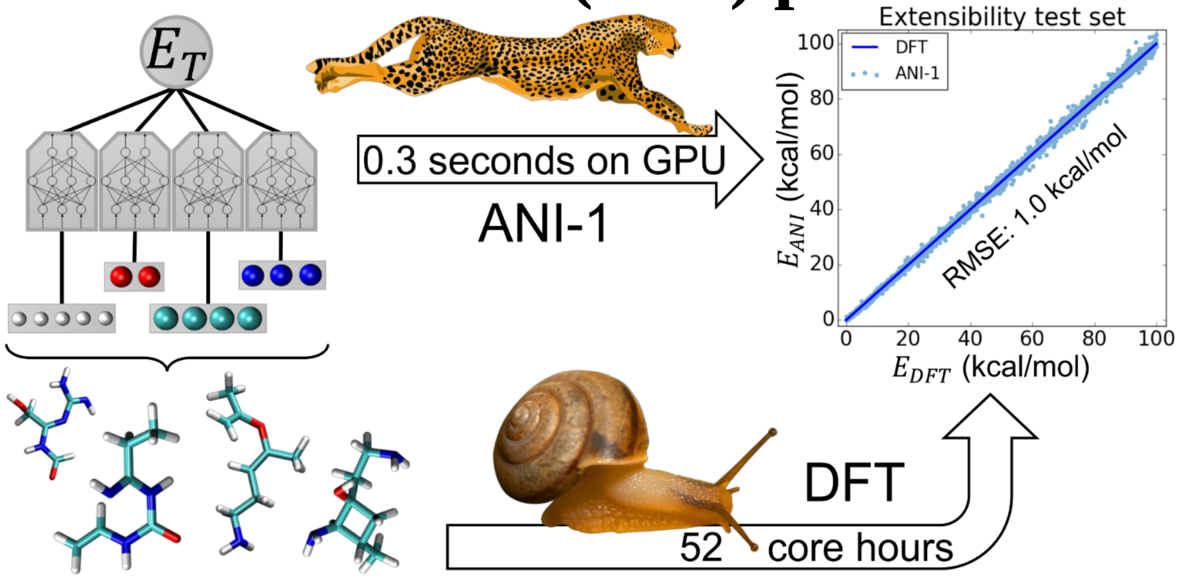
Deep Learning

neural networks (NN)



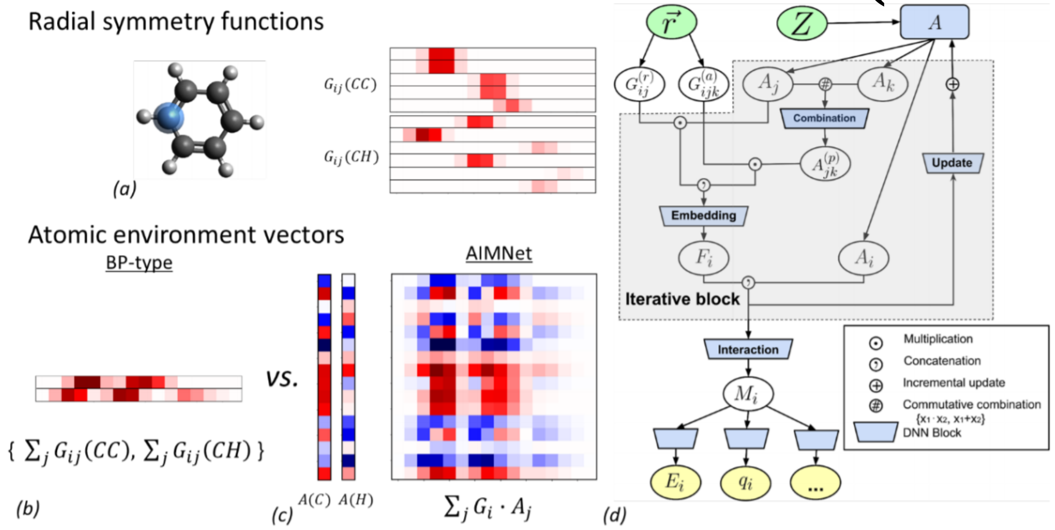
Our work on developing ML potentials

ANAKIN-ME (ANI) potentials



JS Smith, O Isayev, AE Roitberg, *Chem. Sci.*, 2017, **8**, 3192-3203

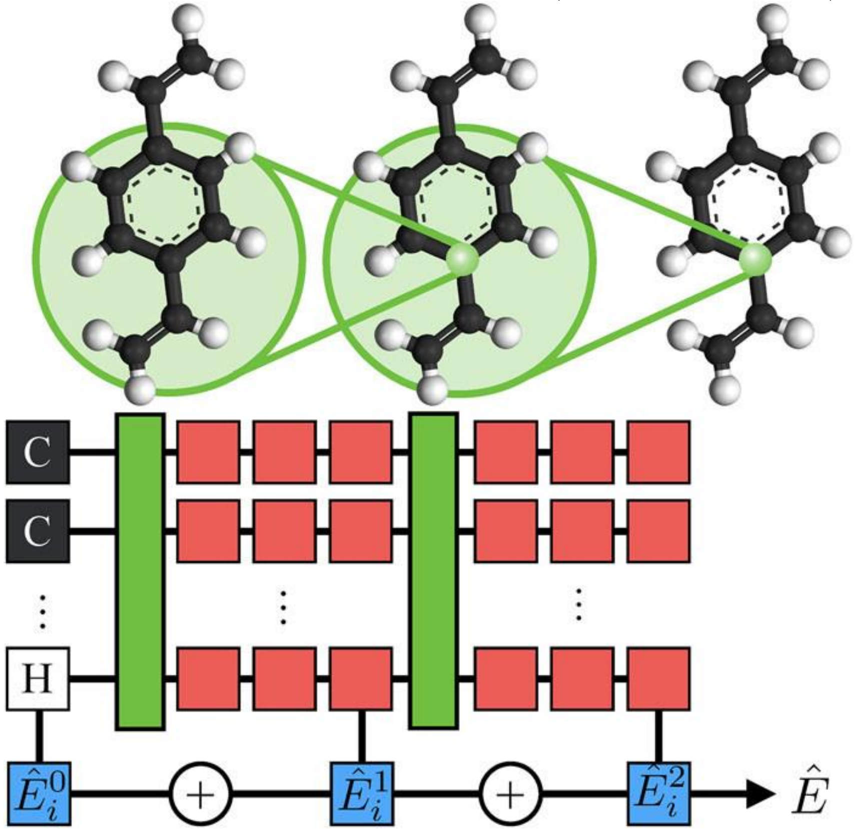
Atoms-in-molecule neural network (AIM-Net)



R Zubatyuk, JS Smith, J Leszczynski, O Isayev

https://doi.org/10.26434/chemrxiv.7151435.v2 2018

Hierarchal Interacting Particle Neural Networks (HIP-NN)



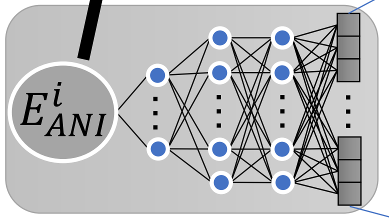
N Lubbers, JS Smith, K Barros

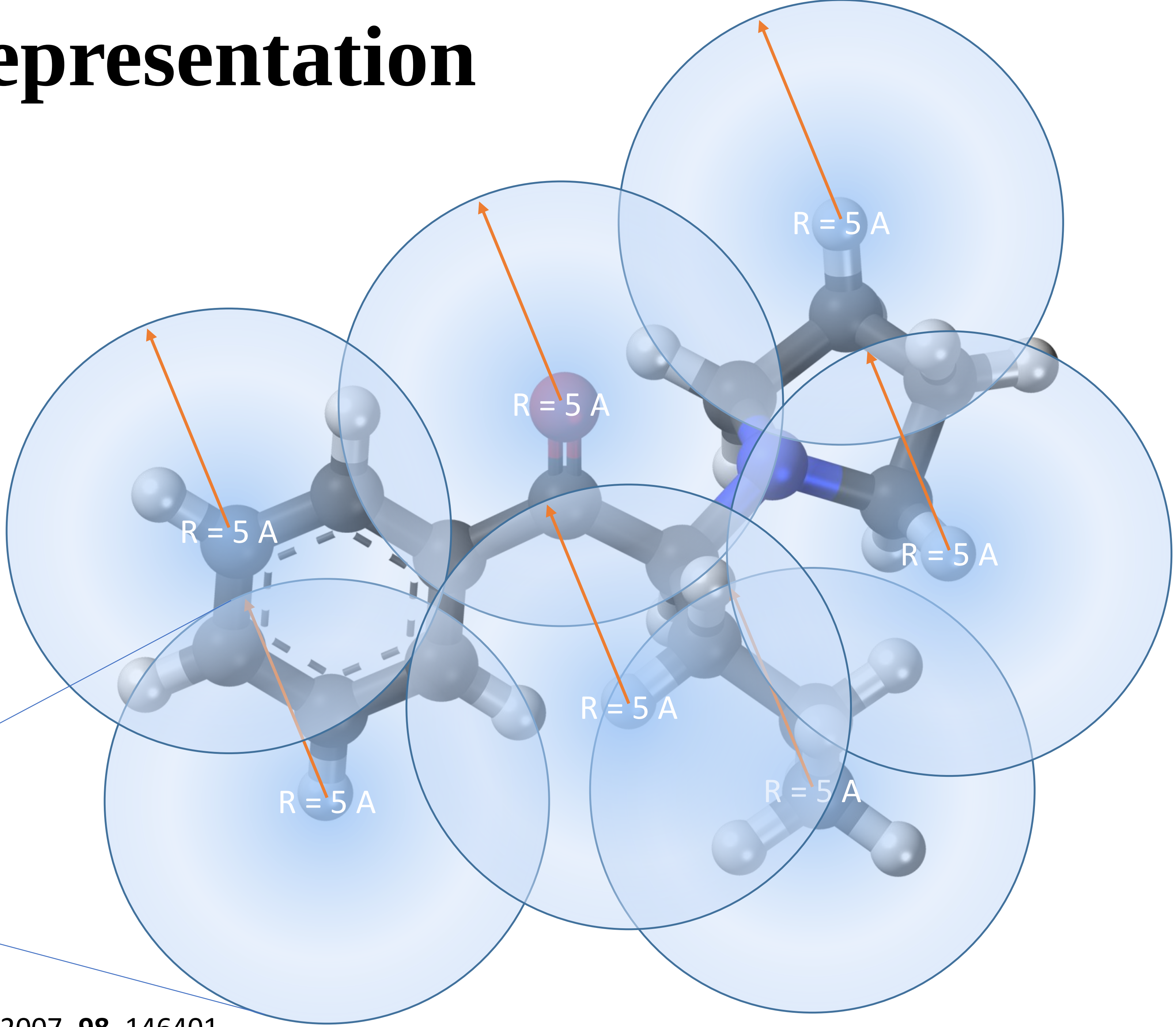
J. Chem. Phys., 2018, **148**, 241715

Molecular Representation

Energy is given by a sum over atomic contributions

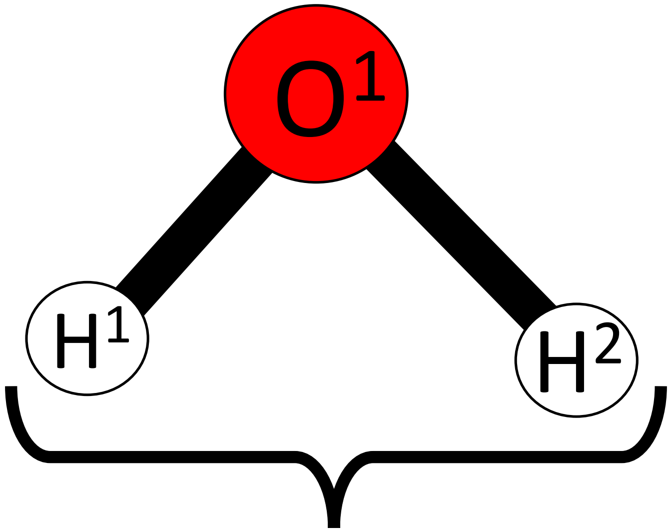
$$E = \sum_{i}^{N} E_i$$





J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.

Descriptors for the ANI ML-based potential



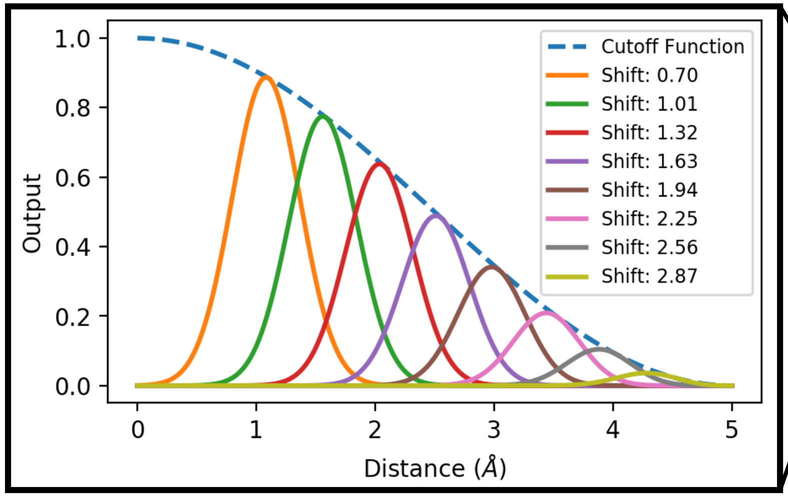
O^1	H^1	H^2
D(H^1)	D(H^2)	D(H^1)
D(H^2)	D(O^1)	D(O^1)
A(H^1;H^2)	A(O^1;H^2)	A(O^1;H^1)

D = distance to atom

A = Angel between atoms

Radial Descriptors

$$G_m^R = \sum_{j \neq i}^{\text{All Atoms}} e^{-\eta(R_{ij}-R_S)^2} f_C(R_{ij})$$

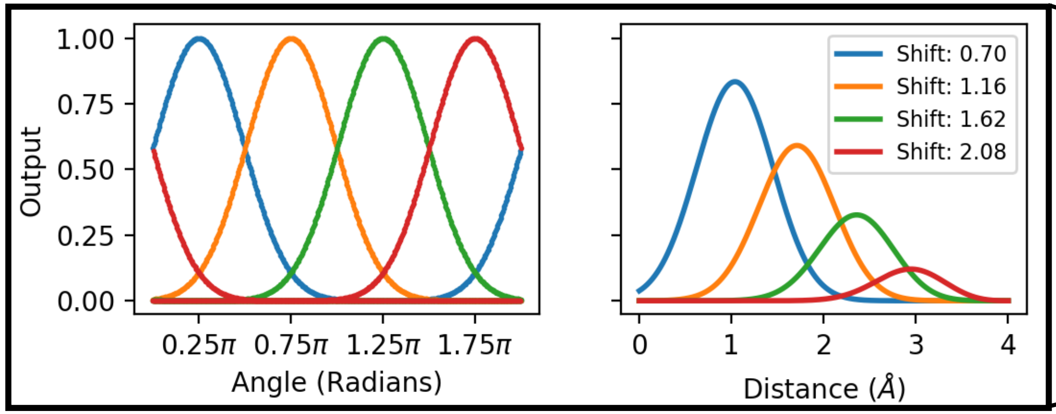


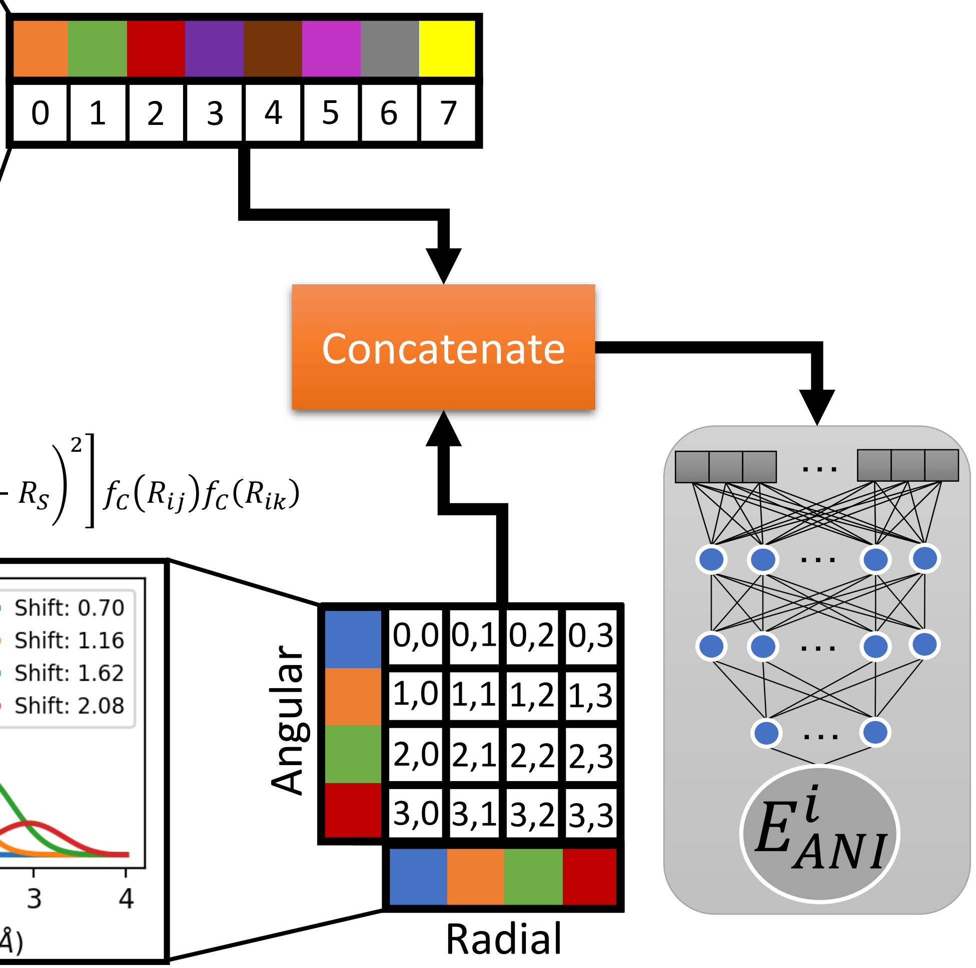
Cutoff function

$$f_C(R_{ij}) = \begin{cases} 0.5 \times \cos\left(\frac{\pi R_{ij}}{R_C}\right) + 0.5 & \text{for } R_{ij} \leq R_C \\ 0.0 & \text{for } R_{ij} > R_C \end{cases}$$

Angular descriptors

$$G_m^{A\text{mod}} = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{All Atoms}} \left(1 + \cos(\theta_{ijk} - \theta_s)\right)^{\zeta} exp\left[-\eta\left(\frac{R_{ij}^2 + R_{ik}^2}{2} - R_S\right)^2\right] f_C(R_{ij}) f_C(R_{ik})$$

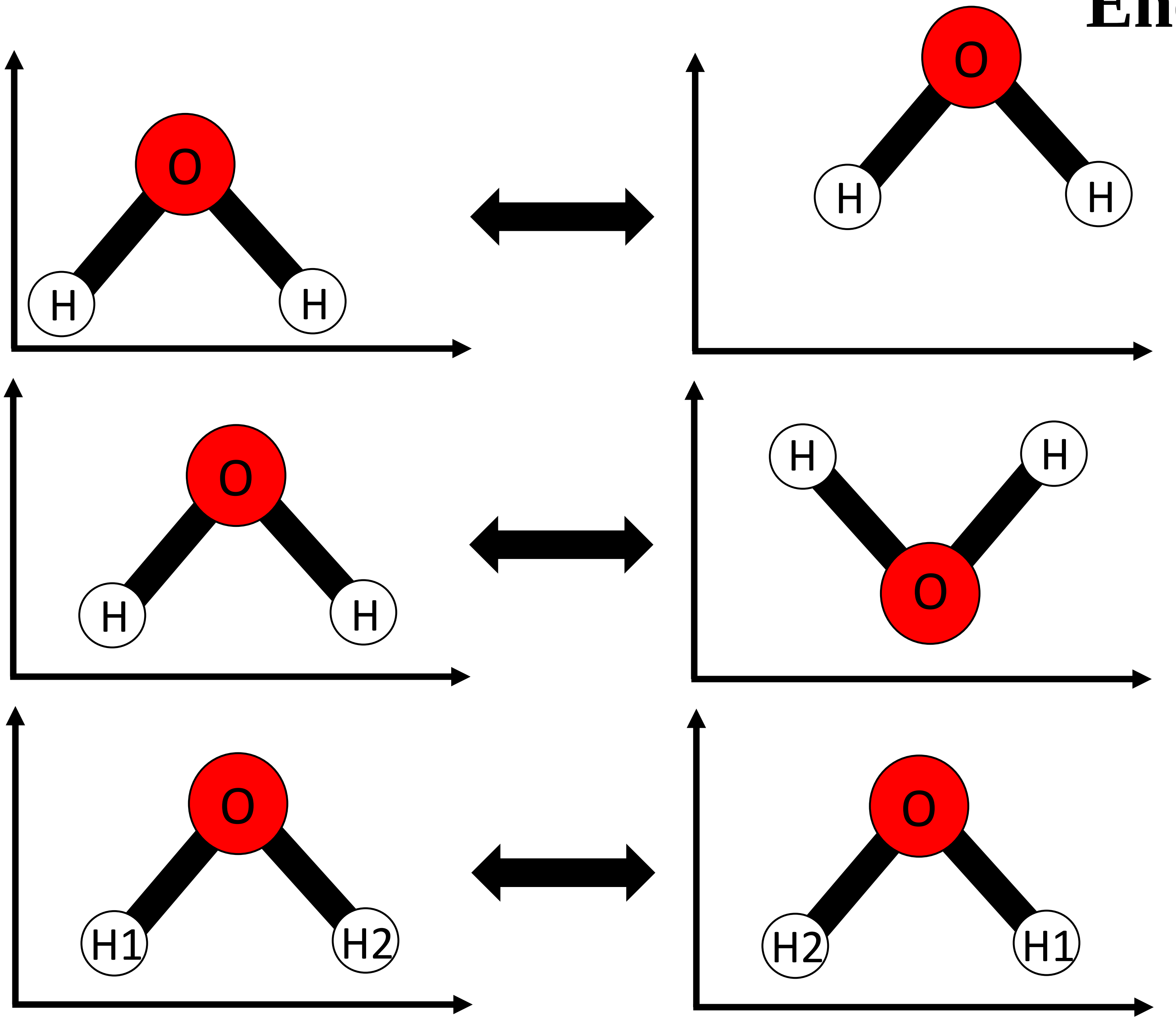




JS Smith, O Isayev, AE Roitberg, *Chem. Sci.*, 2017, **8**, 3192-3203

J Behler and M Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401

Energy must be invariant with respect to:

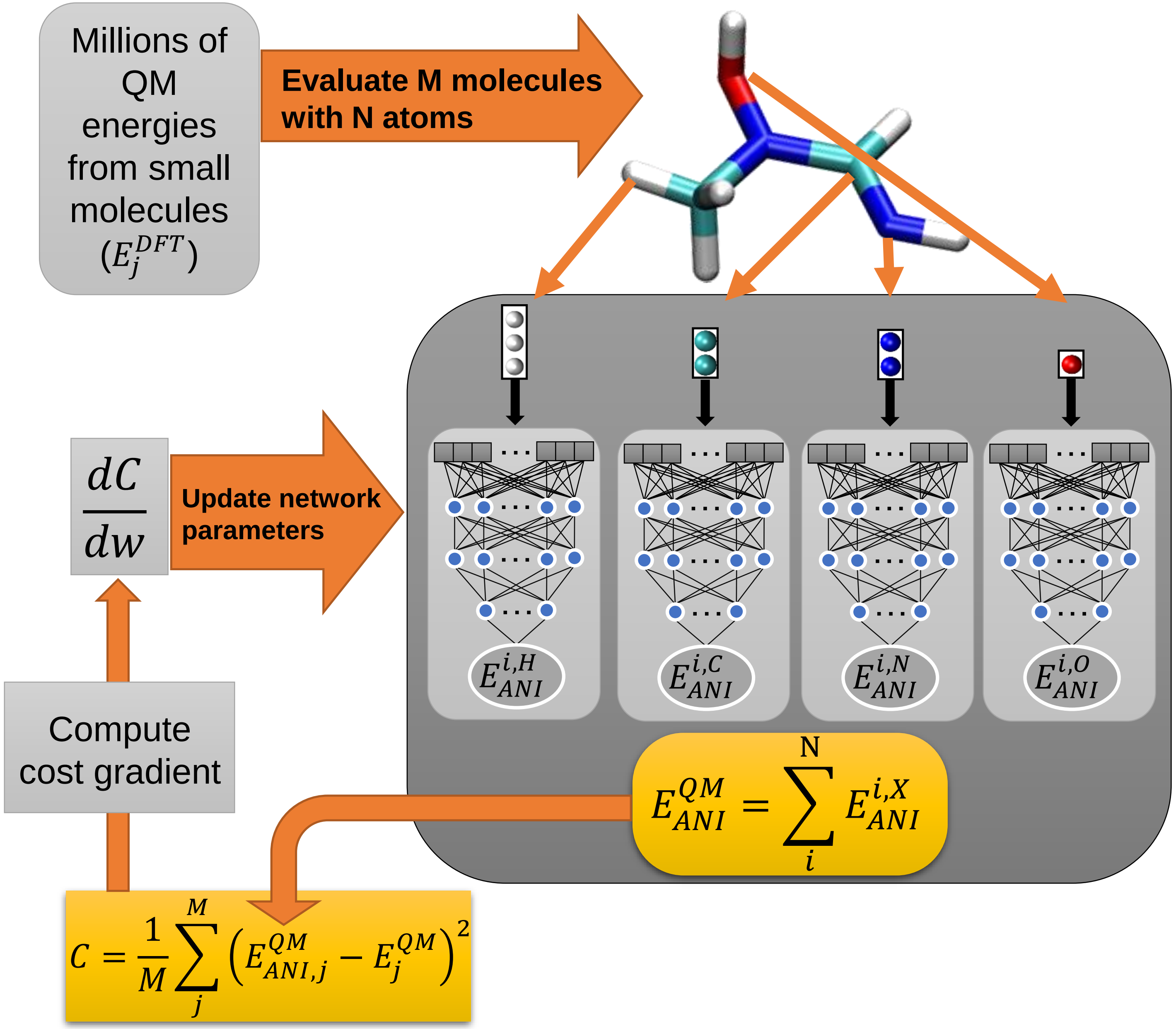


Translation

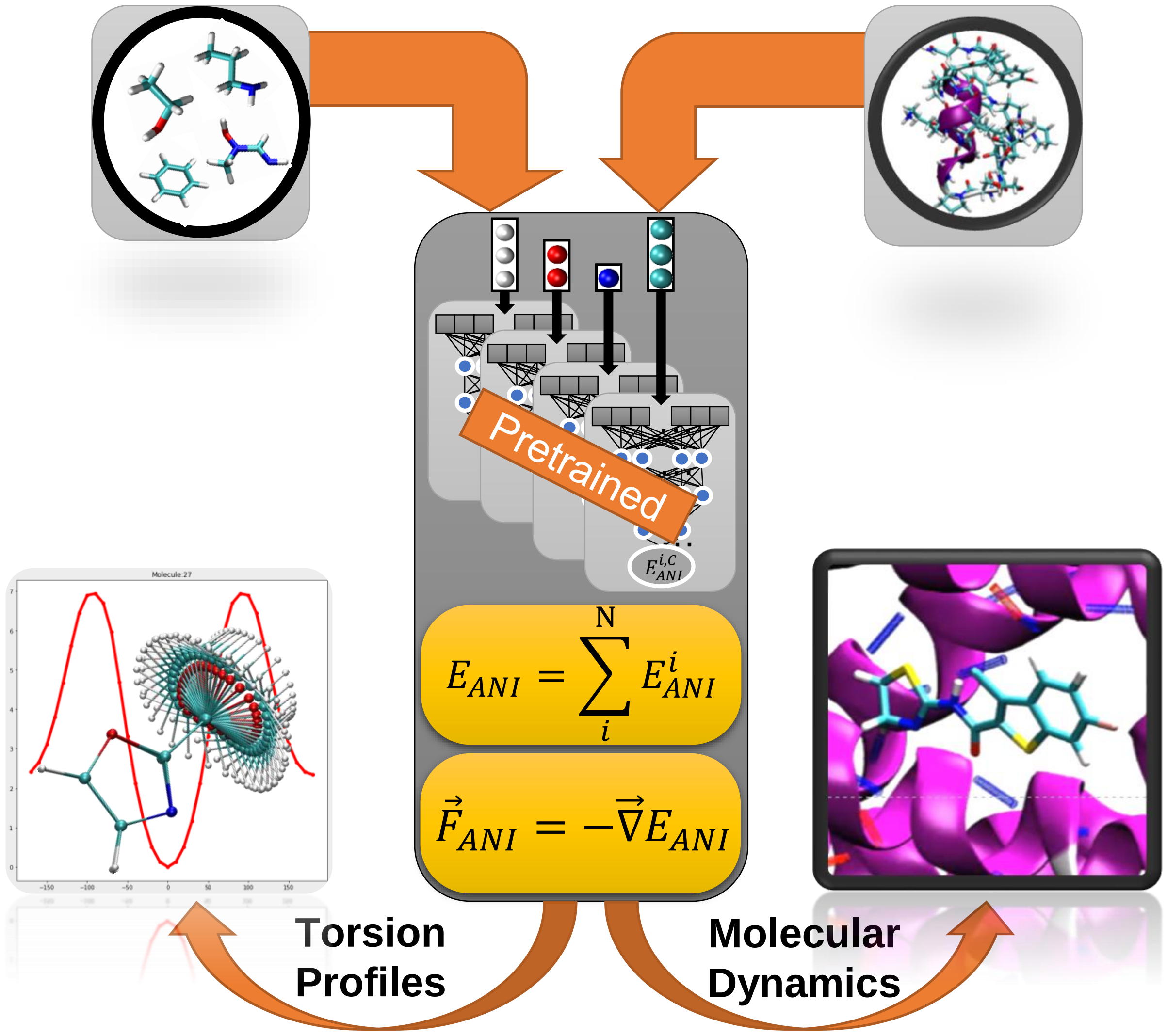
Rotation

Permutation

ANI potential training

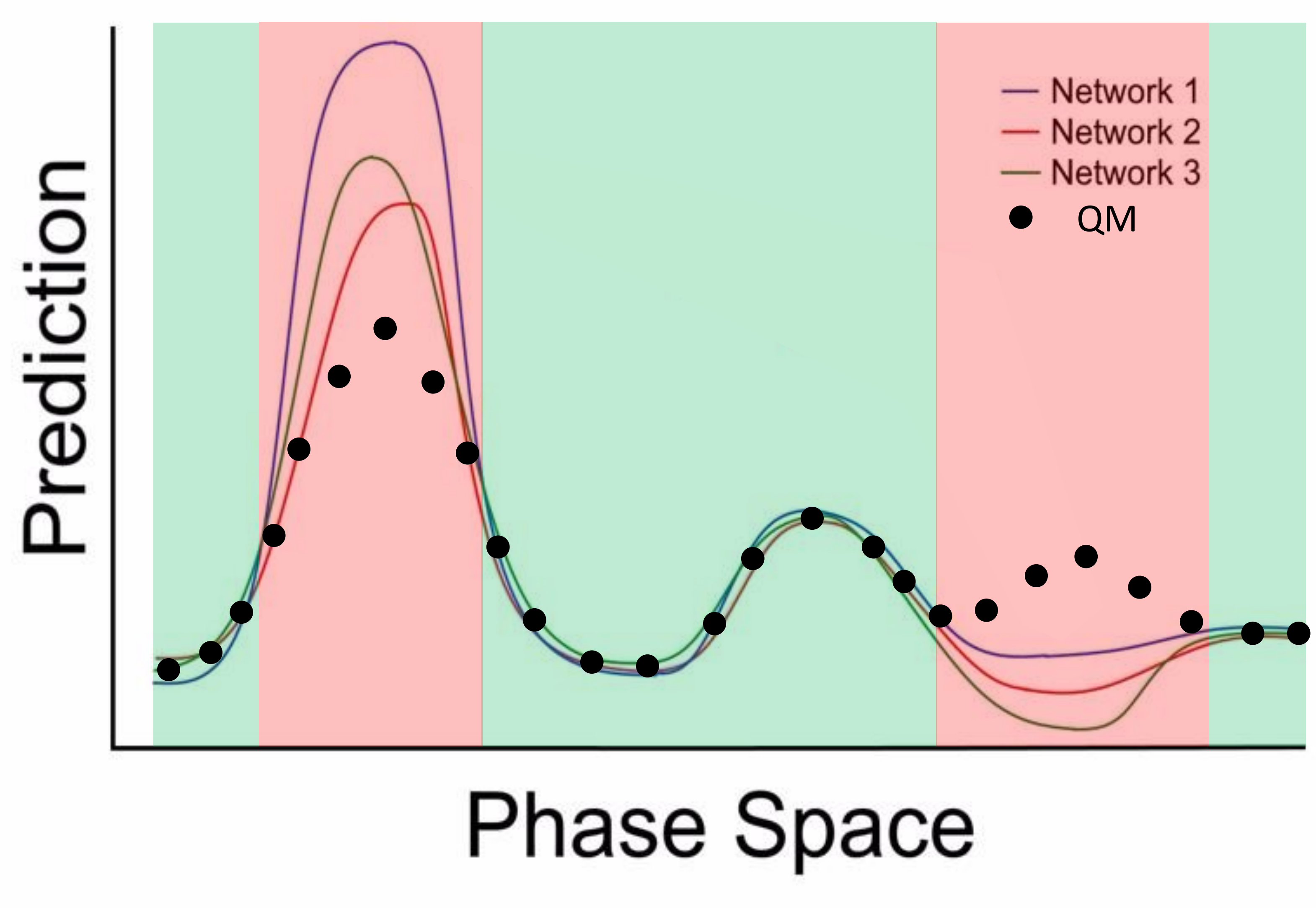


ANI potential application



Can we predict when the model is wrong?

Ensemble disagreement can drive data generation



Good data coverage

Bad data coverage

Active Learning - The Big Picture

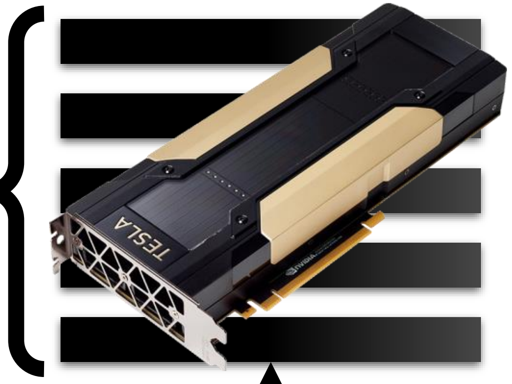
An automated and self-consistent data generation framework

Non-equilibrium Conformational sampler

New test data

Check ensemble disagreement

Ensemble of ANI networks



Train network ensemble

Molecule Sampling
(e.g. GDB small molecule database, small peptides, drug like molecules)

Structure Pools

Compute Cluster

Computations with QM

ANI-1x Dataset
(i.e. energies, forces, dipoles)

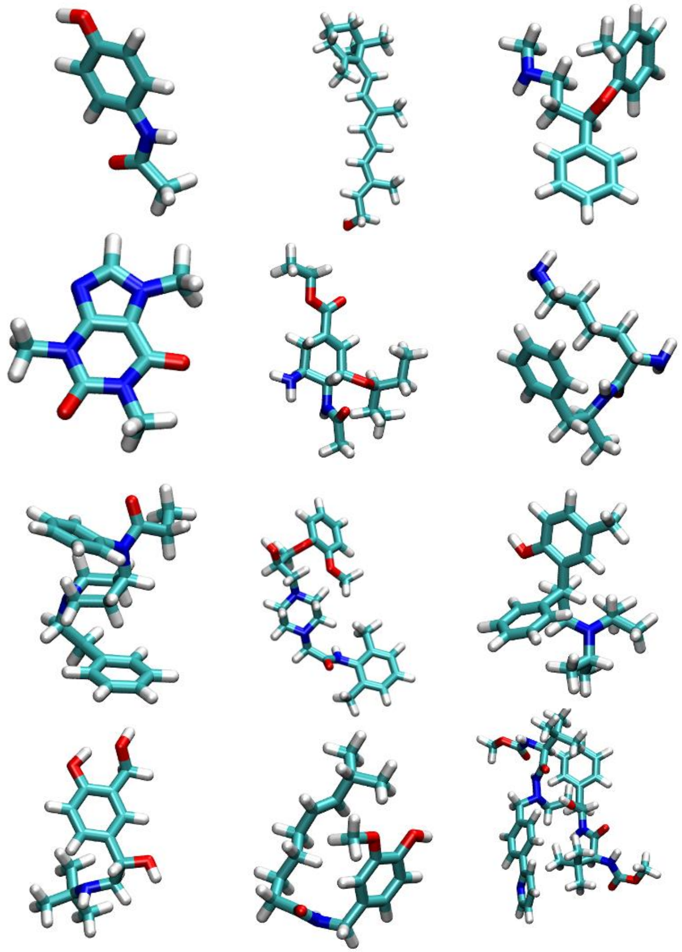
JS Smith, et al.; *The Journal of Chemical Physics*, (**2018**), 148 (24), 241733

Testing transferability and extensibility

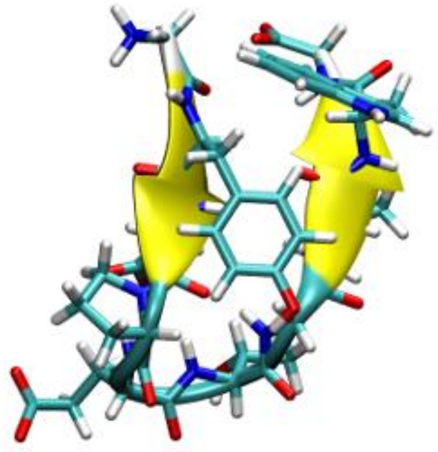
ANI-MD Benchmark

128 frames from 1ns trajectories @ 300K for each:

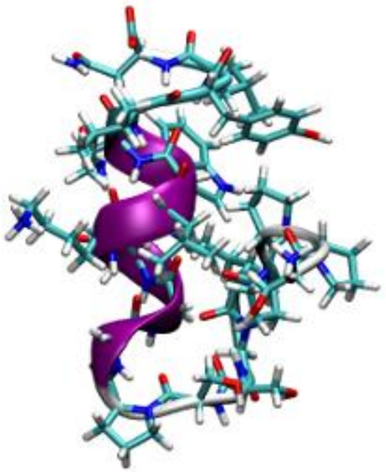
Various drug molecule



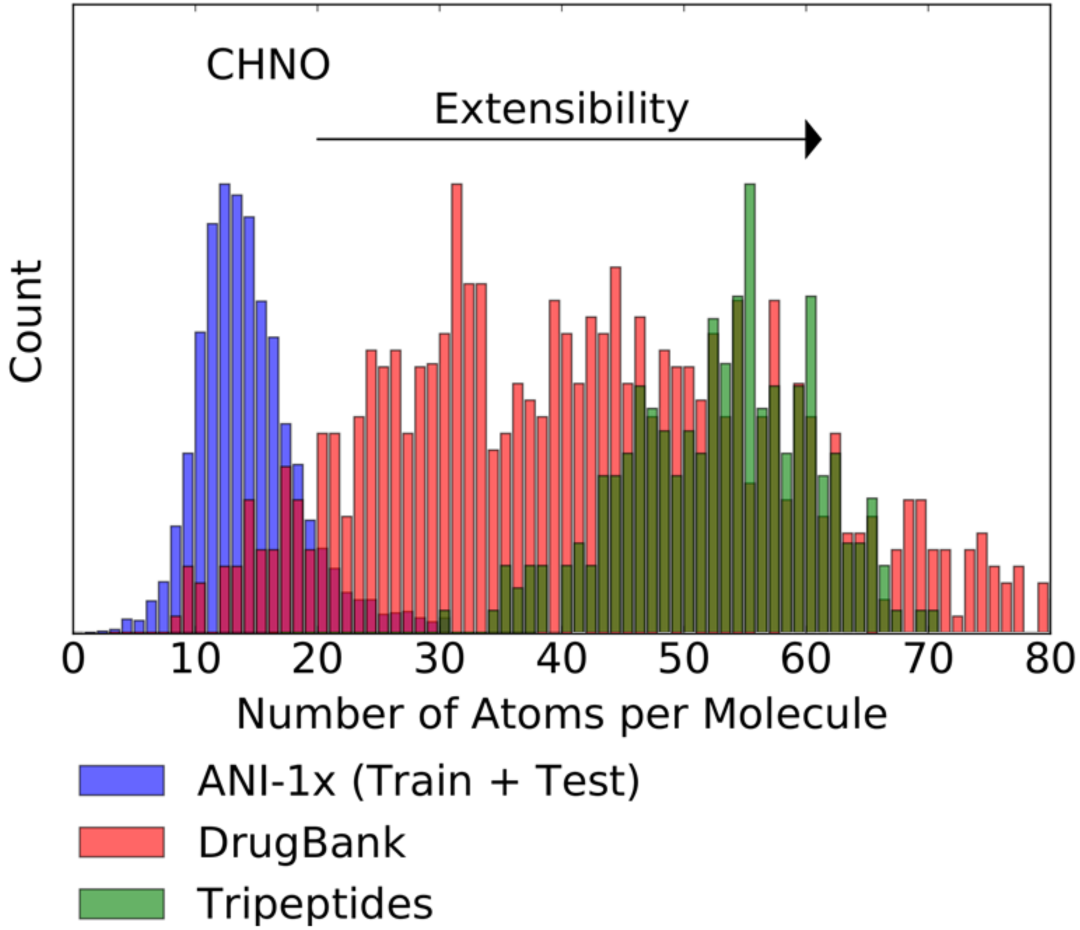
Chignolin



Trp-cage



DrugBank and Tripeptide Benchmarks



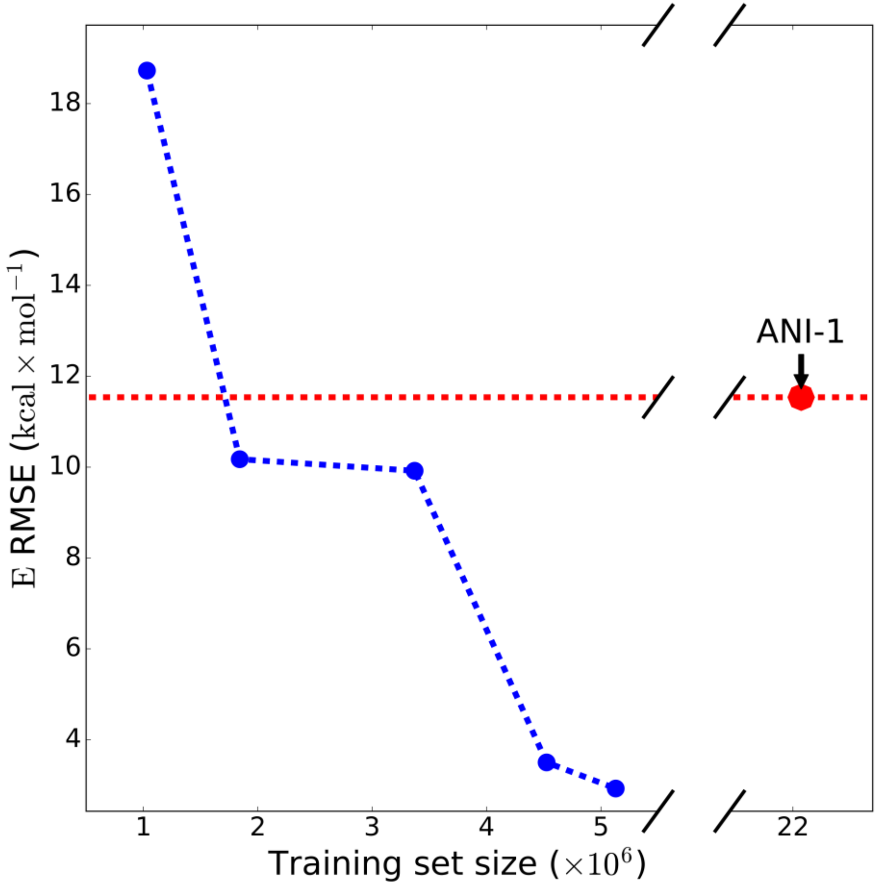
JS Smith, et al.; *The Journal of Chemical Physics*, (**2018**), 148 (24), 241733

Active-learning results vs. random sampling

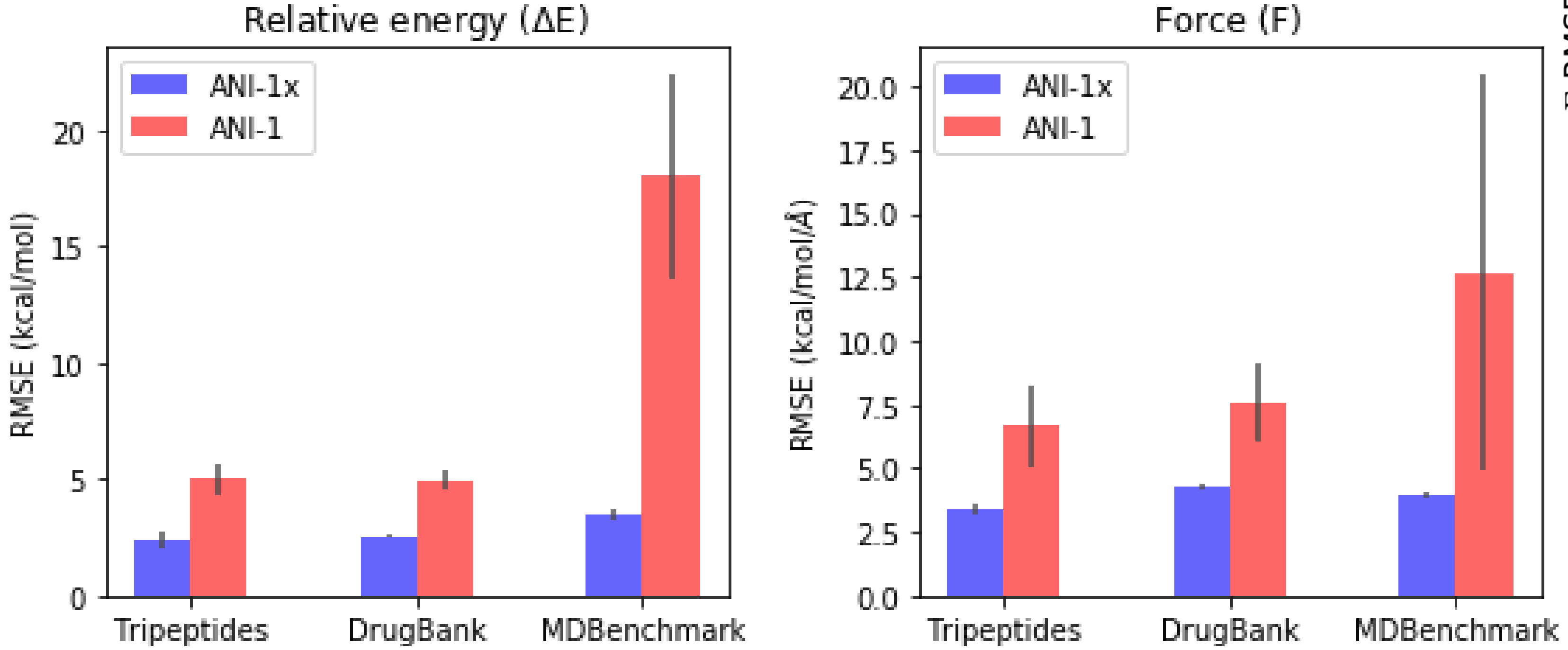
Dataset size comparison

	ANI-1	ANI-1x
Datapoints	22M	5M

Active learning progression



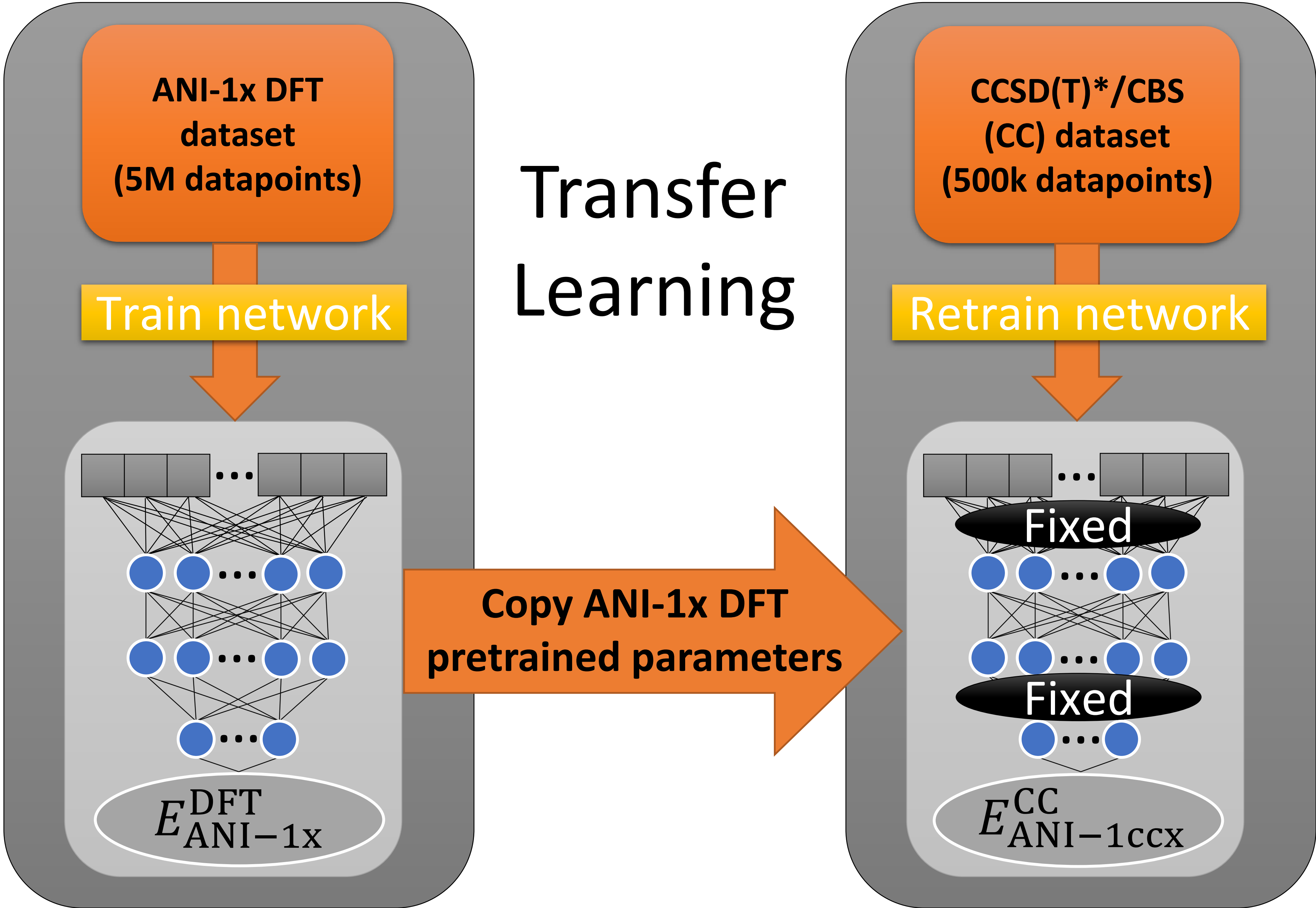
Relative E and F RMSE comparison



JS Smith, et al.; *The Journal of Chemical Physics*, (**2018**), 148 (24), 241733

Transferring knowledge from DFT to CCSD(T)

- Subsample 10% of ANI-1x training data (0.5M of 5M)
- Recompute CCSD(T)/CBS level
- 340k parameters fixed, re-train 60k
- 10^7 faster than DFT

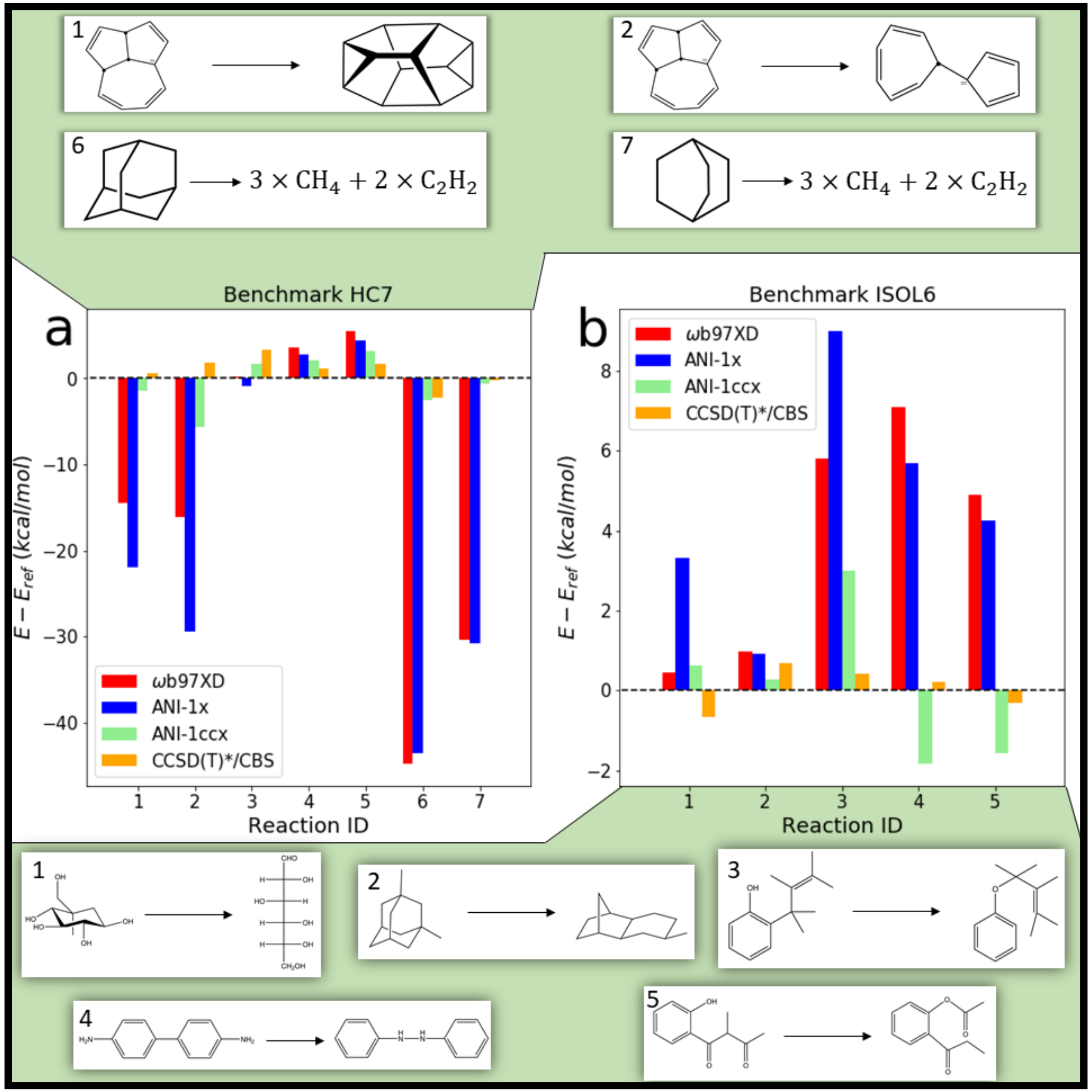


JS Smith, et al.
https://doi.org/10.26434/chemrxiv.6744440.v1
(2018)

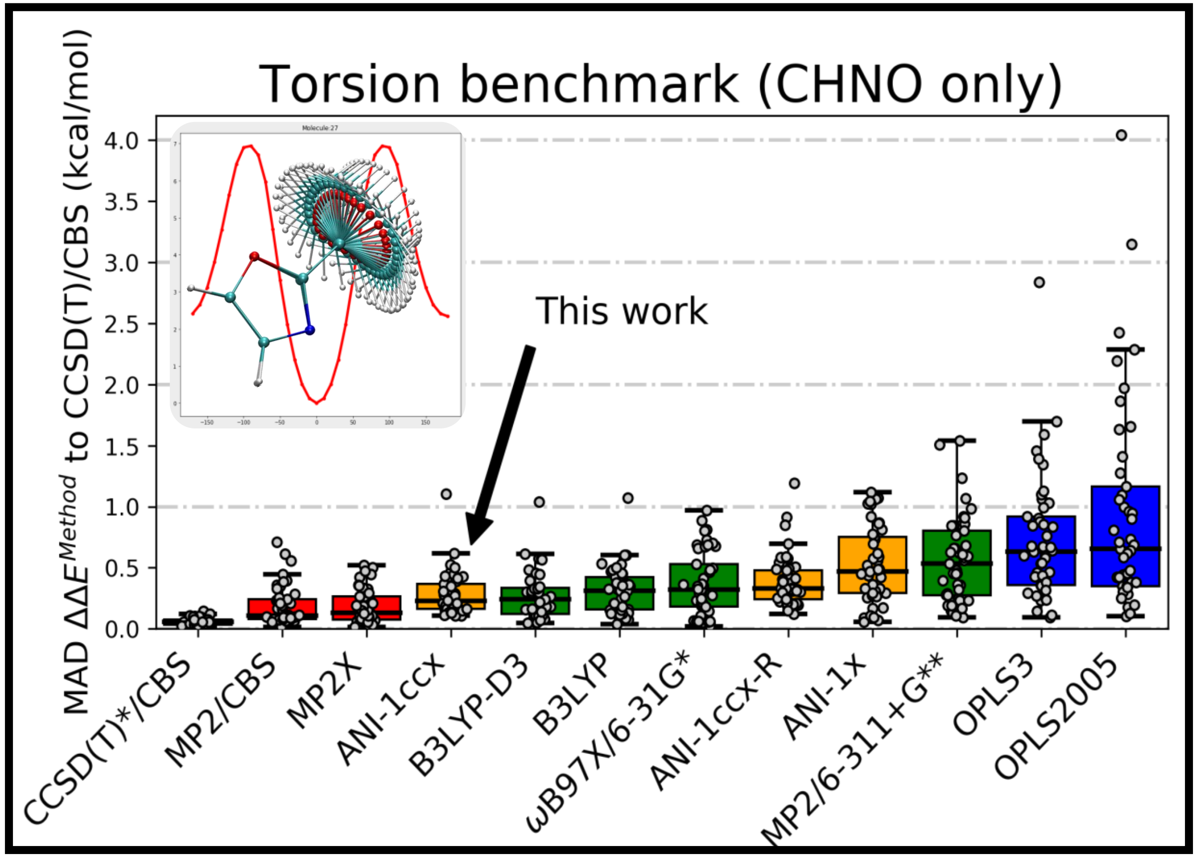
Outsmarting Quantum Chemistry Through Transfer Learning

JS Smith, B Nebgen, R Zubatyuk, N Lubbers, C Devereux, K Barros, S Tretiak, O Isayev, A Roitberg

https://doi.org/10.26434/chemrxiv.6744440.v1 **2018** (under review at Nat. Comm.)



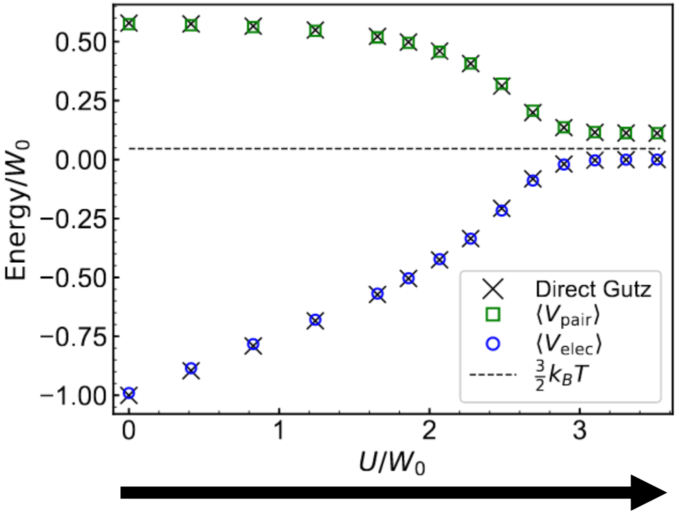
- New ANI-1ccx model outperforms DFT on reaction energies and torsional profiles
- A 24 core hours calculation for CCSD(T)/CBS takes 2 GPU microseconds for ANI-1ccx



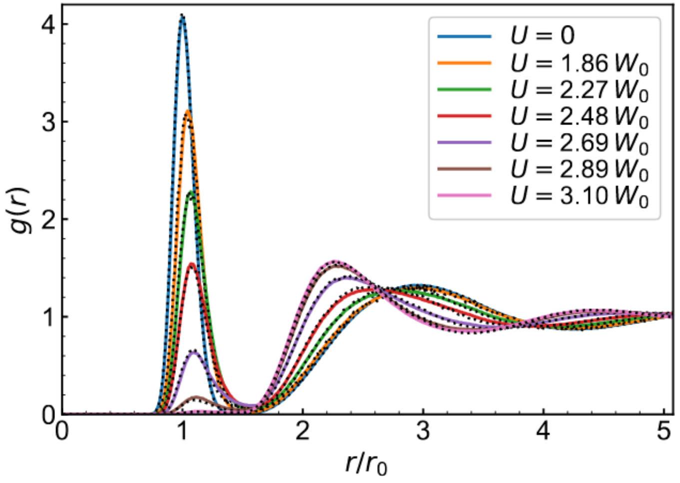
Machine learning for molecular dynamics with strongly correlated electrons

Hidemaro Suwa, Justin S. Smith, Nicholas Lubbers, Cristian D. Batista, Gia-Wei Chern, Kipton Barros

https://arxiv.org/abs/1811.01914 **2018** (under review at Phys. Rev. Lett.)

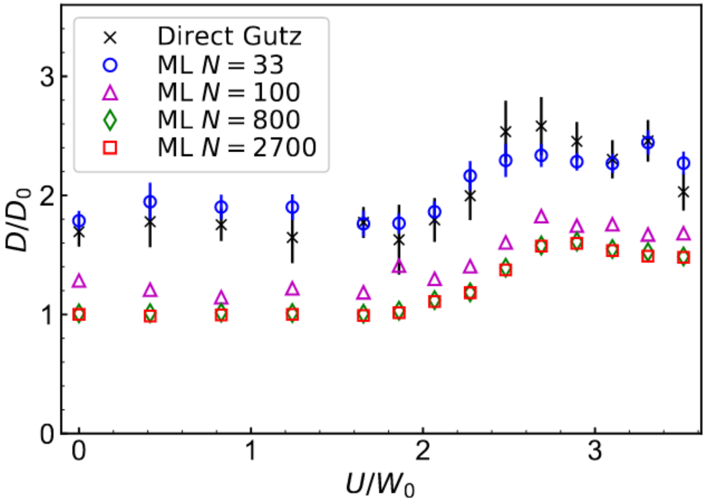


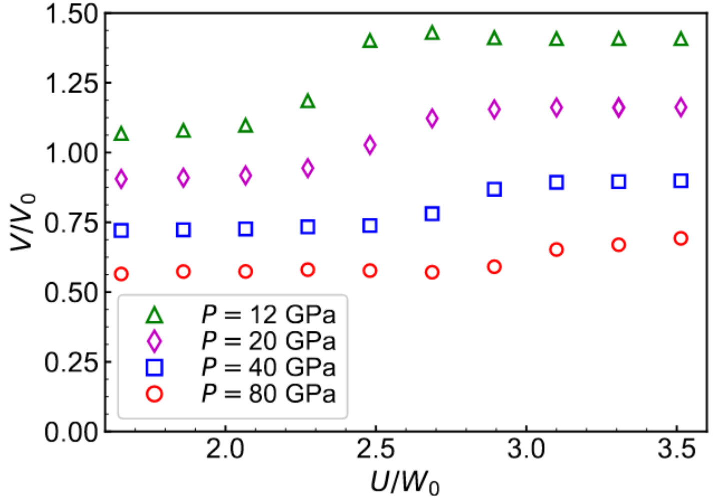
Metal Insulator



Trained an ML model to a toy system with variable Hubbard U through the Gutzwiller approximation

Accurately reproduces a Mott transition on systems of 2700 atoms.





ML metal potentials with active learning!

Our approach: minimize use of expert knowledge for **maximum generality**

Los Alamos Team

Benjamin Nebgen – T-1
Kipton Barros – T-1
Saryu Fensin – MST-8
Tim Germann – T-1
Leonid Burakovsky – T-1
Nicholas Lubbers – CCS-3
Sergei Tretiak – T-1

Energy $E[\mathbf{r}_1, \mathbf{r}_2, \ldots]$

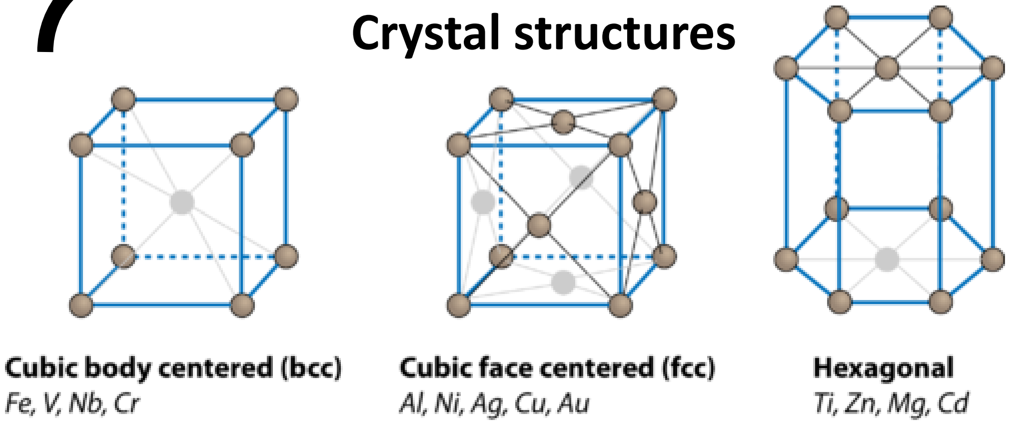
Force $\mathbf{f}_i = -\nabla_i E$

Applications

$$m\frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{f}_i$$

ML Potential

Crystal structures



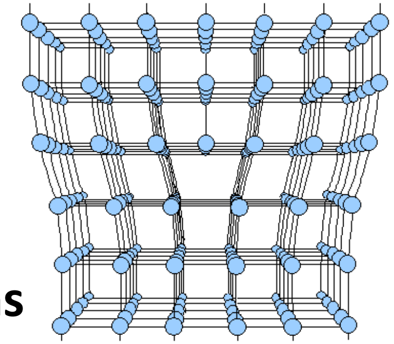
https://courses.lumenlearning.com/cheminter/chapter/crystal-structures-of-metals/

Melts

Amorphous solid

Extreme conditions

Defects



The General AL Framework for ML potentials

A codebase for active learning on large GPU clusters (e.g. Sierra or Summit)

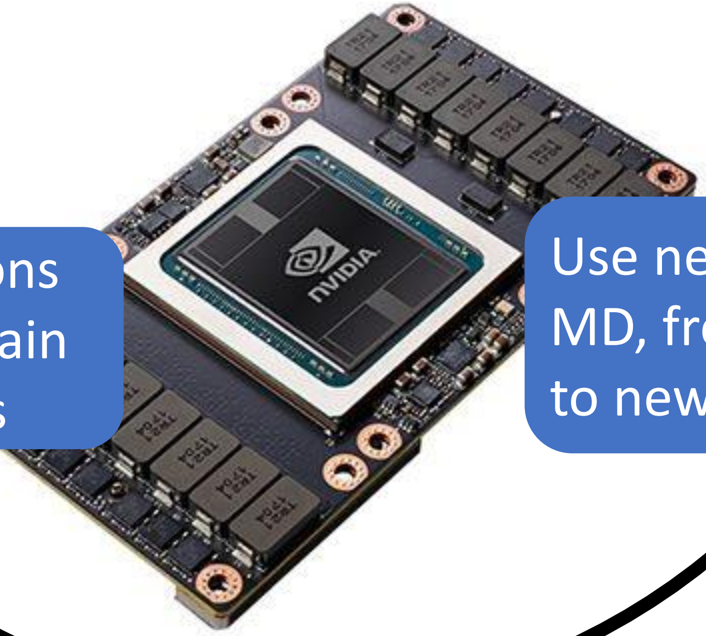
Design Principles:

- Fully autonomous
- Interchangeable QM
- Interchangeable ML
- Assortment of built-in sampling methods
- Built-in testing suite
- Capable of scaling to 1000s of nodes

Retrain ML models on all available DFT data

Use new potentials to run MD, freezing from liquid to new solid structures

Run new DFT calculations whenever ML is uncertain about force predictions



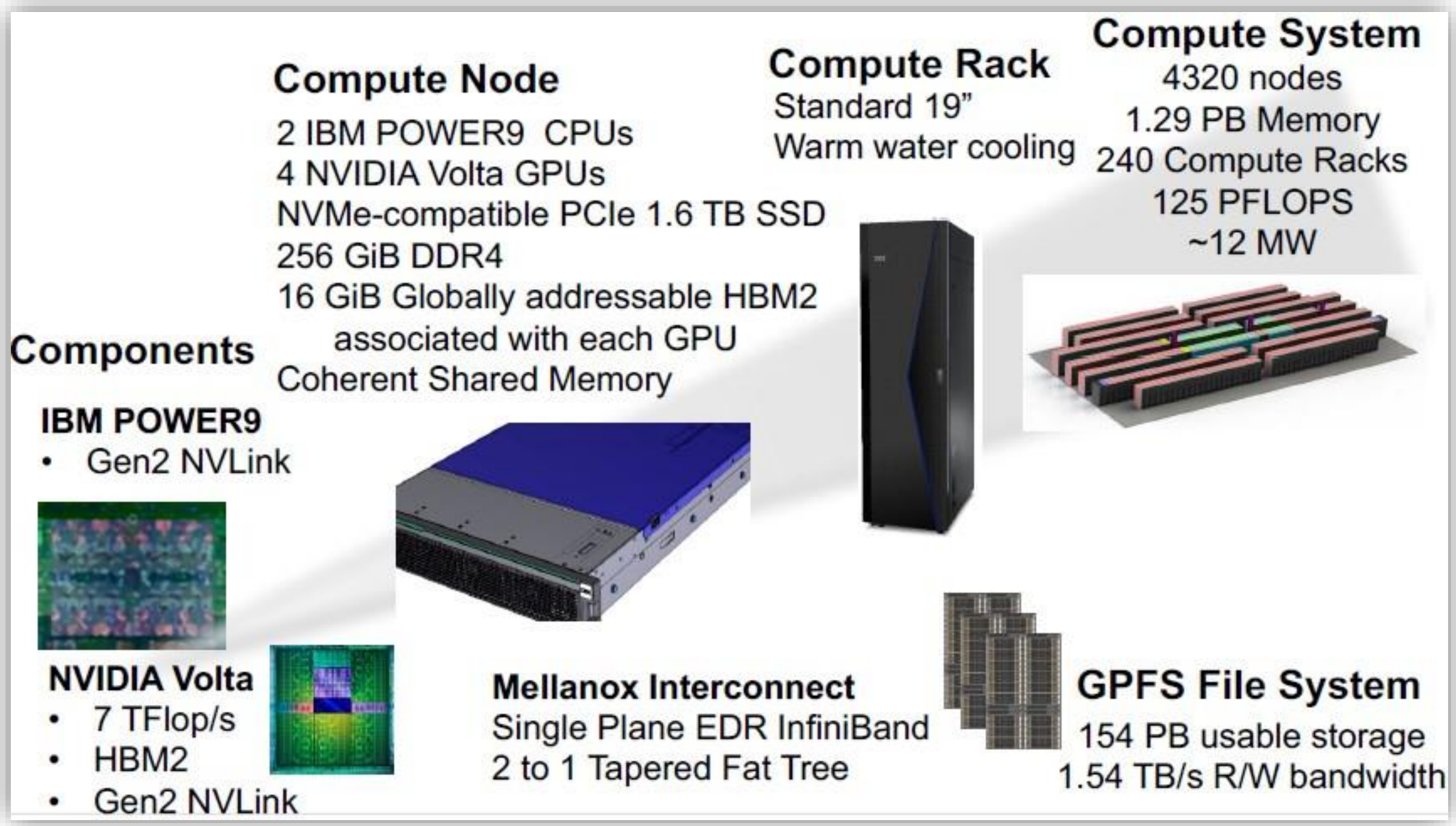
Kipton Barros

Benjamin Nebgen

Application of active learning to build ML potentials

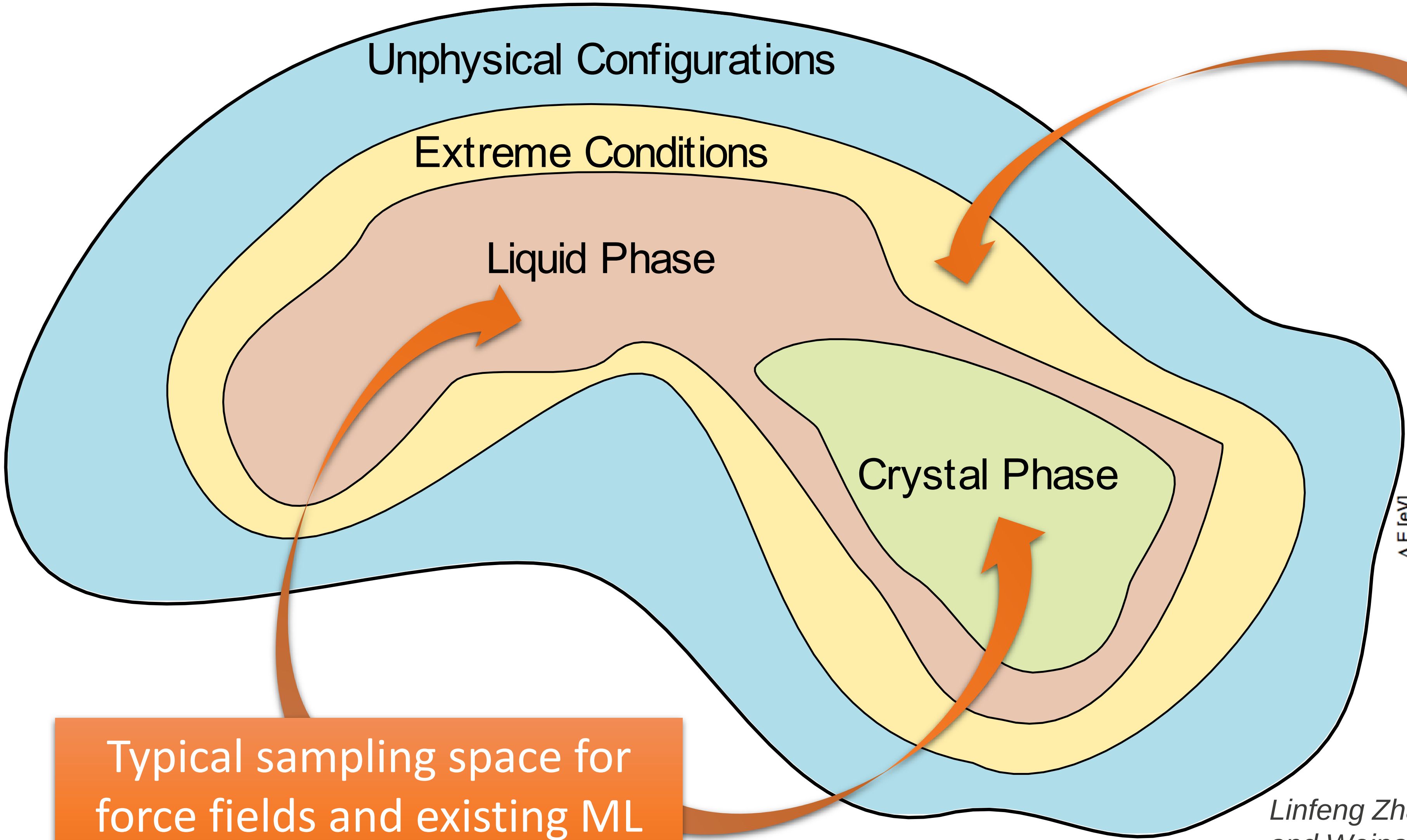
- Build a general active learning framework
- Framework interface with Quantum Espresso (QE) for DFT
- In 2 months we ran 10-20k DFT calculations on systems with 50-200 metal atoms
- Elements explored **Al**, **Sn**, Ga, Cu
- We are still evaluating results

Open science access on LLNL's Sierra super computer



How should we sample to build a general model?

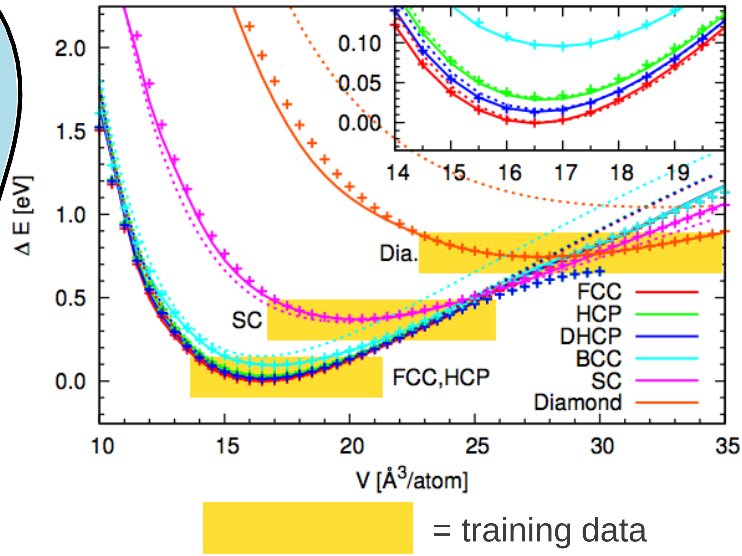
All possible configurations for a metal



Where extreme conditions and rare events exist (e.g. shock simulations)

Typical sampling space for force fields and existing ML potentials

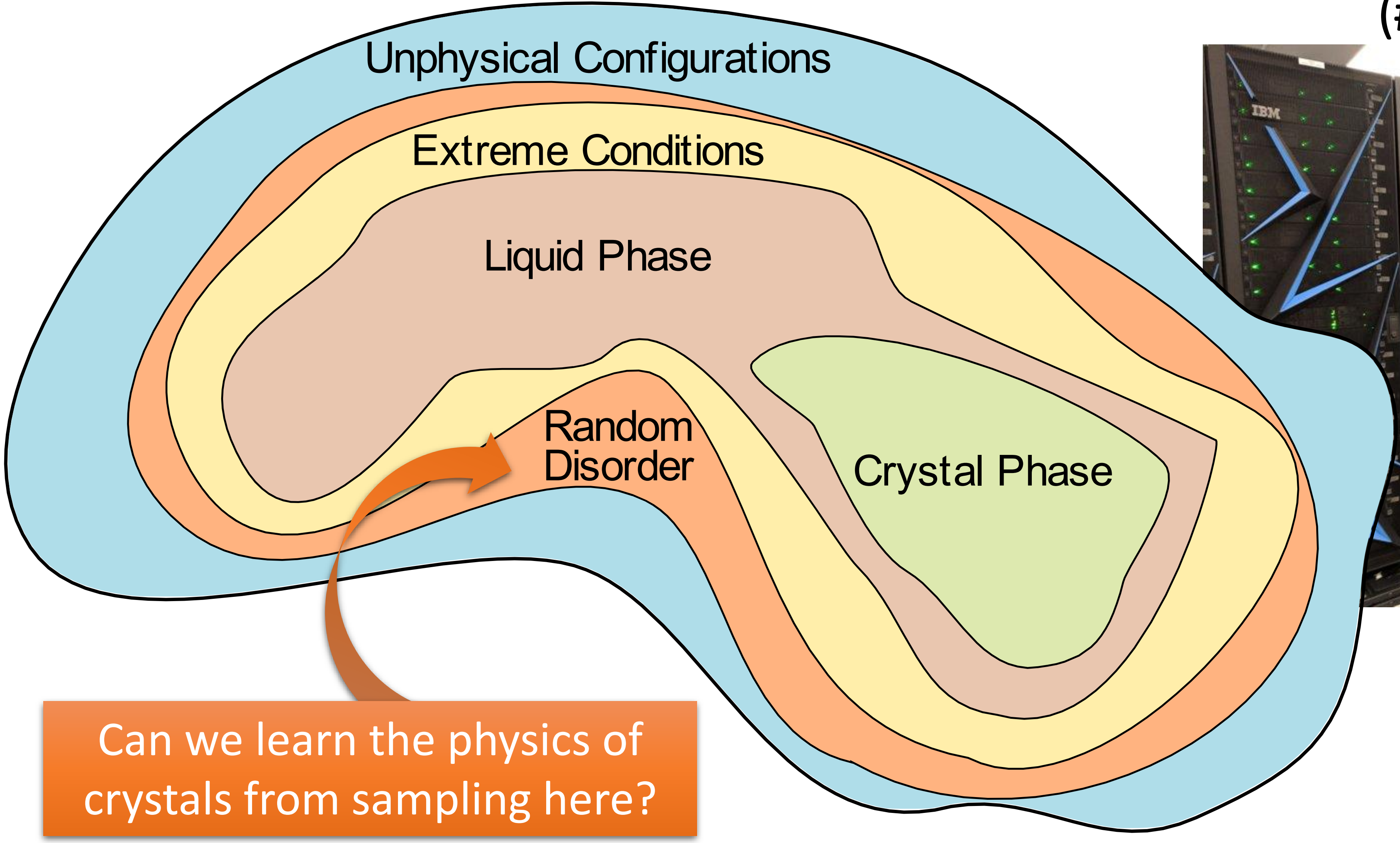
Recent published work



Linfeng Zhang, De-Ye Lin, Han Wang, Roberto Car, and Weinan E, Active Learning of Uniformly Accurate Inter-atomic Potentials for Materials Simulation, [arXiv:1810.11890]

How should we sample to build a general model?

All possible configurations for a metal



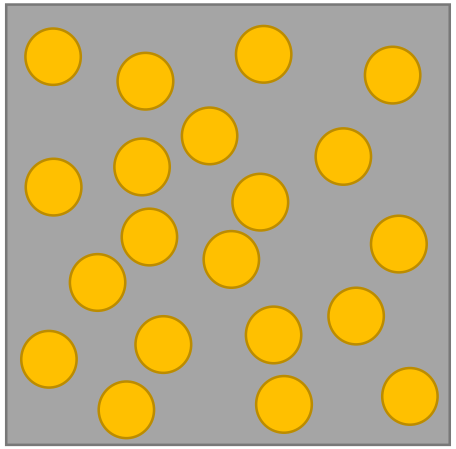
LLNL Sierra
(#2 on TOP500 list)

Thanks for the open science early access allocation!

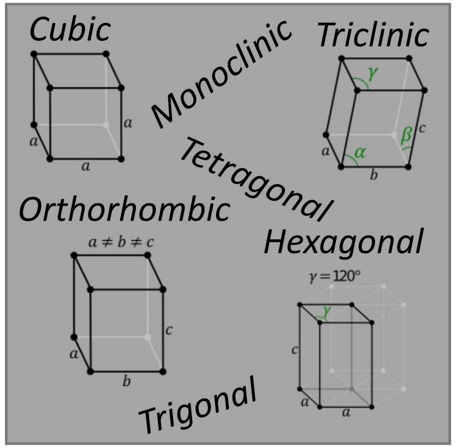
Sampling techniques

Random sampling

Technique 1
Disorder



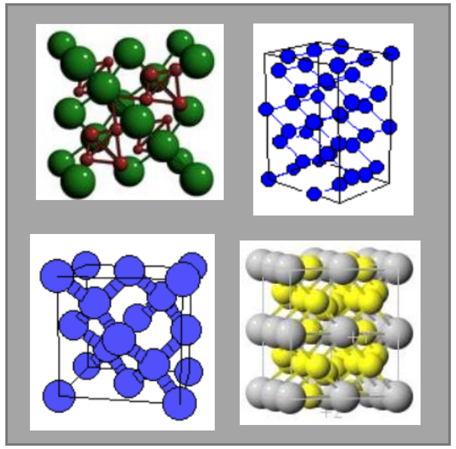
Technique 2
Space group[1]



- Configuration selected by ML
- Minimum atomic distances restrained
- Density kept within a set range
- Random a, b, c lattice constant
- Box size kept minimal

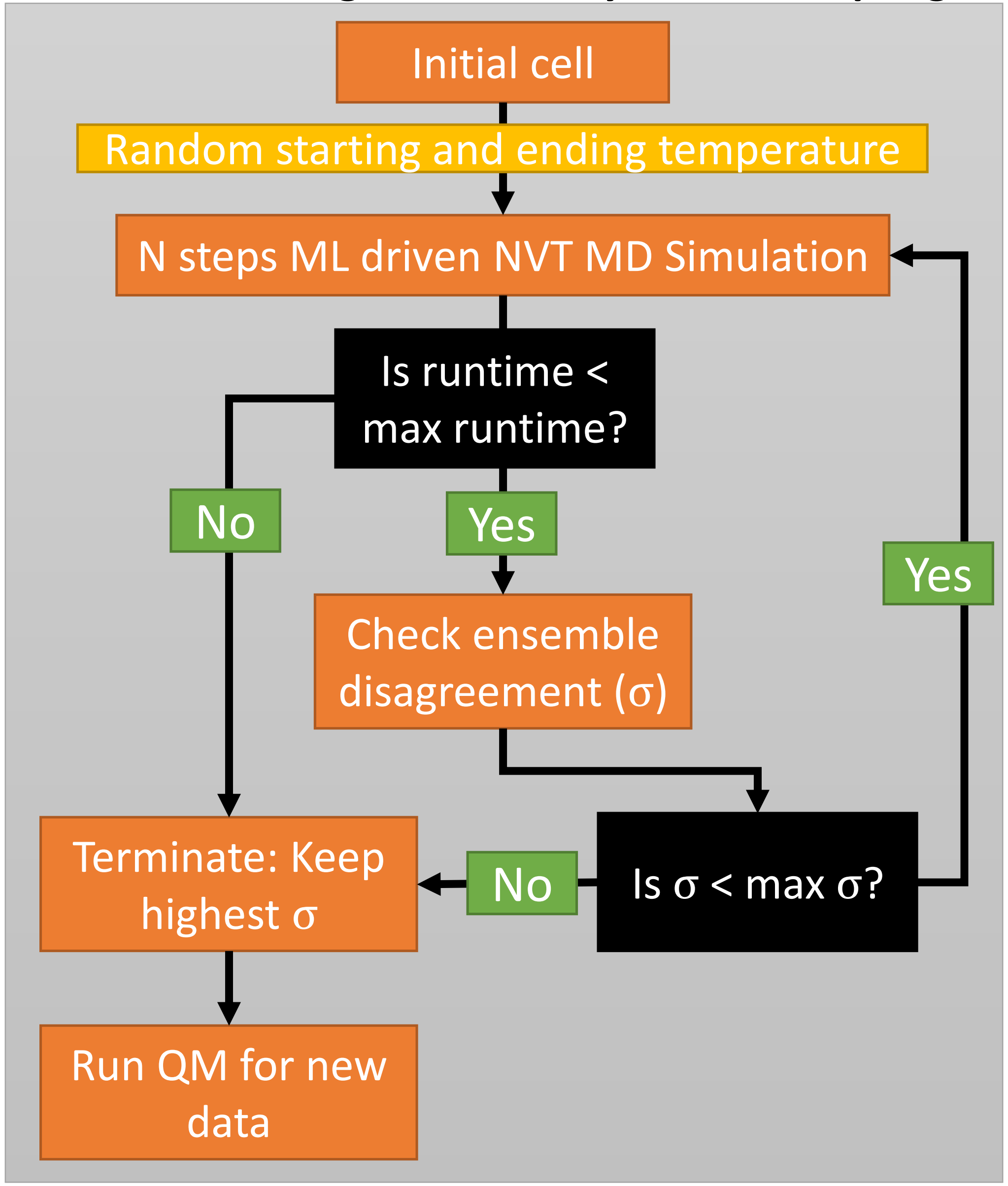
Selected sampling

Technique 3
Crystal[2]



- Crystal selected by human
- Random perturbation by 0.25A
- Random a, b, c lattice constant

Active learning molecular dynamics sampling



1) Images from: https://en.wikipedia.org/wiki/Space_group
2) Images from: https://homepage.univie.ac.at/michael.leitner/lattice/index.html

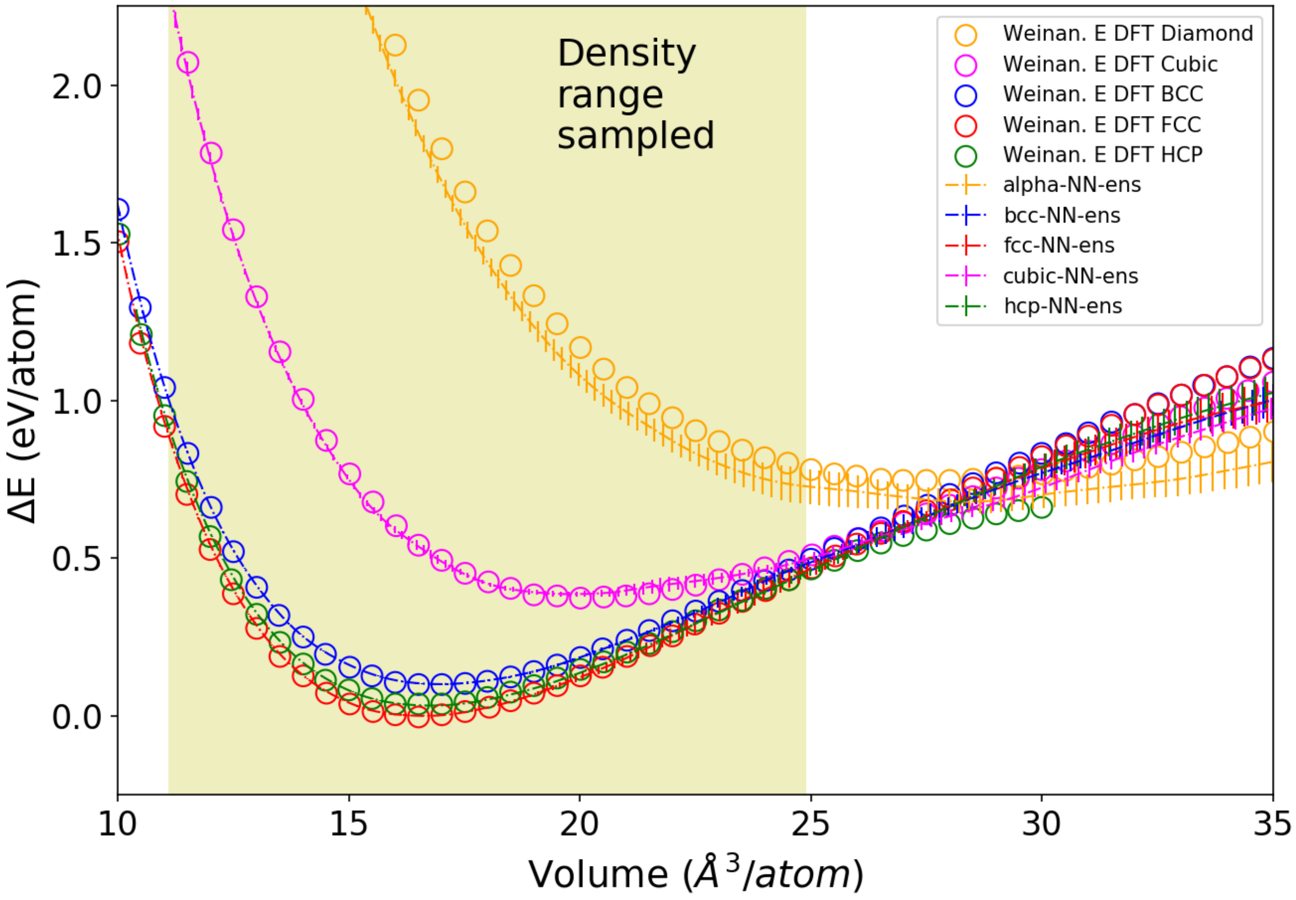
Application on Aluminum (Al)!

DISCLAIMER!

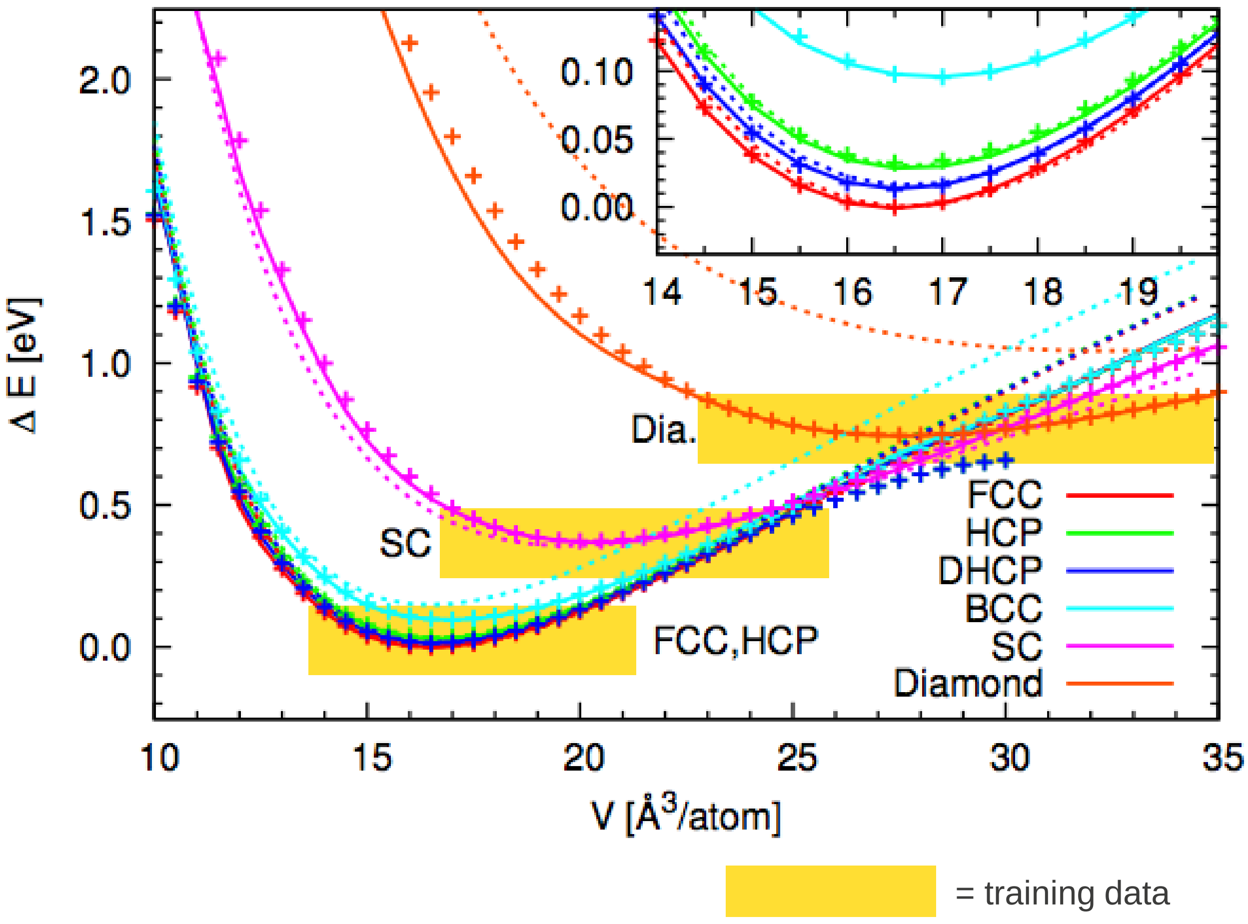
We trained to DFT and are about to show results comparing to experiment.

Select crystal vs. random disorder MD sampling for Al

No human knowledge used in sampling
Sampling technique 1 only: Disorder
(our current work)



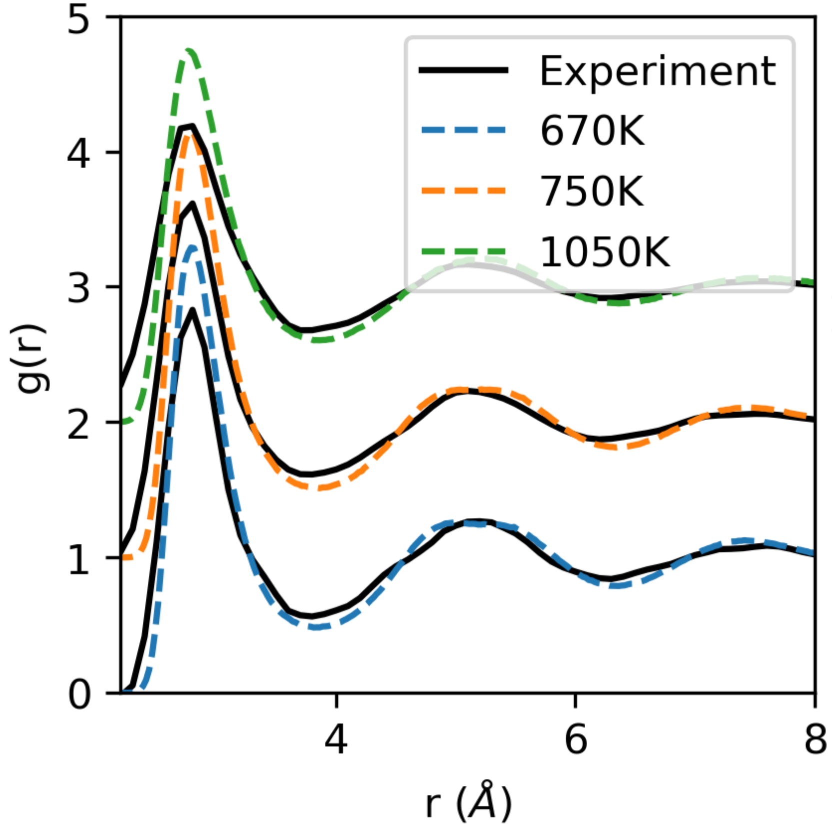
Crystals chosen based on human knowledge (previous literature)

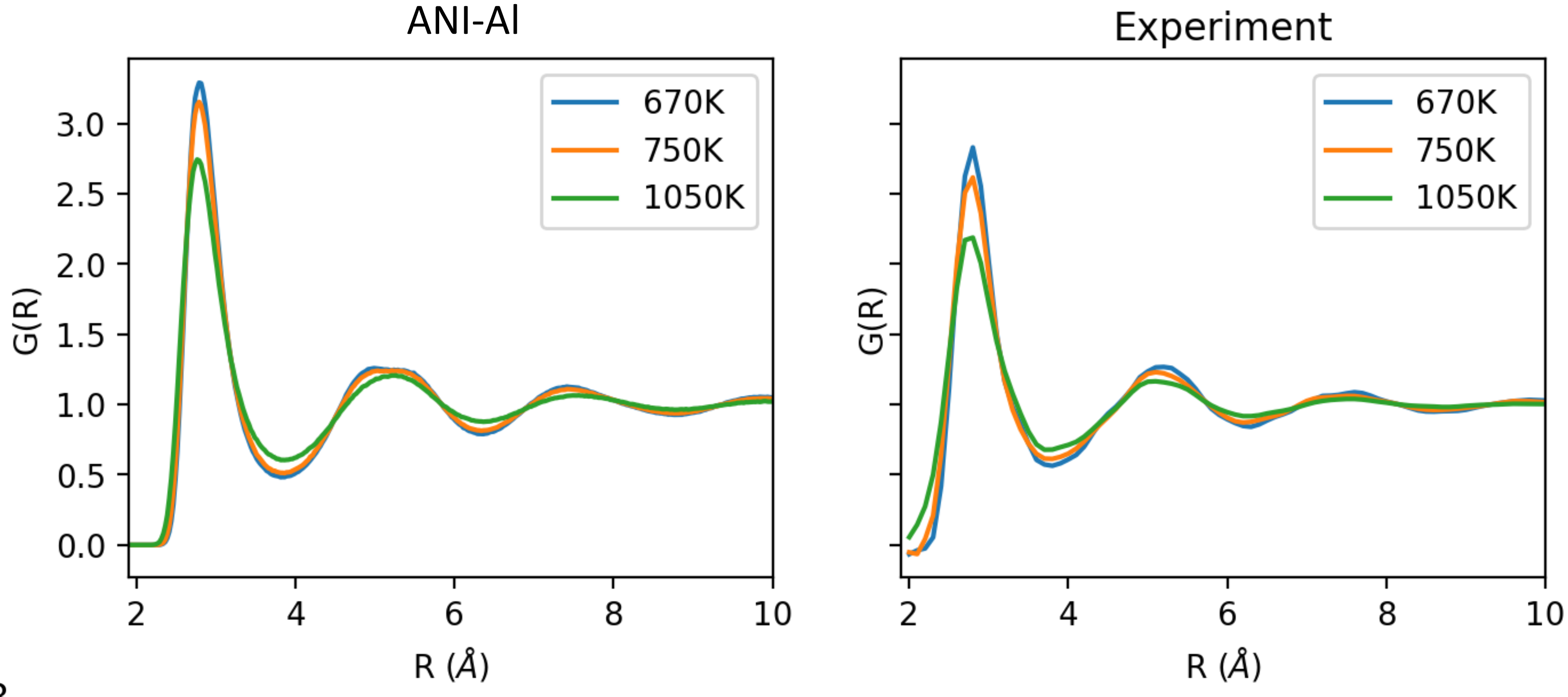


Linfeng Zhang, De-Ye Lin, Han Wang, Roberto Car, and Weinan E, Active Learning of Uniformly Accurate Inter-atomic Potentials for Materials Simulation, [arXiv:1810.11890]

RDF of liquid Al using our ML potential

- 125ps of NPT for equilibration
- 125ps of NVT at equilibrated density
- 2048 atoms system
- Trained to DFT (PBE)
- Density vs exp ~11% off (on the level of typical DFT error)



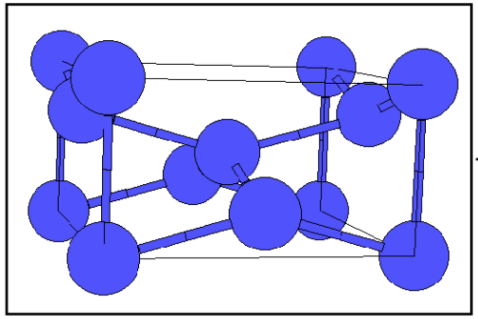


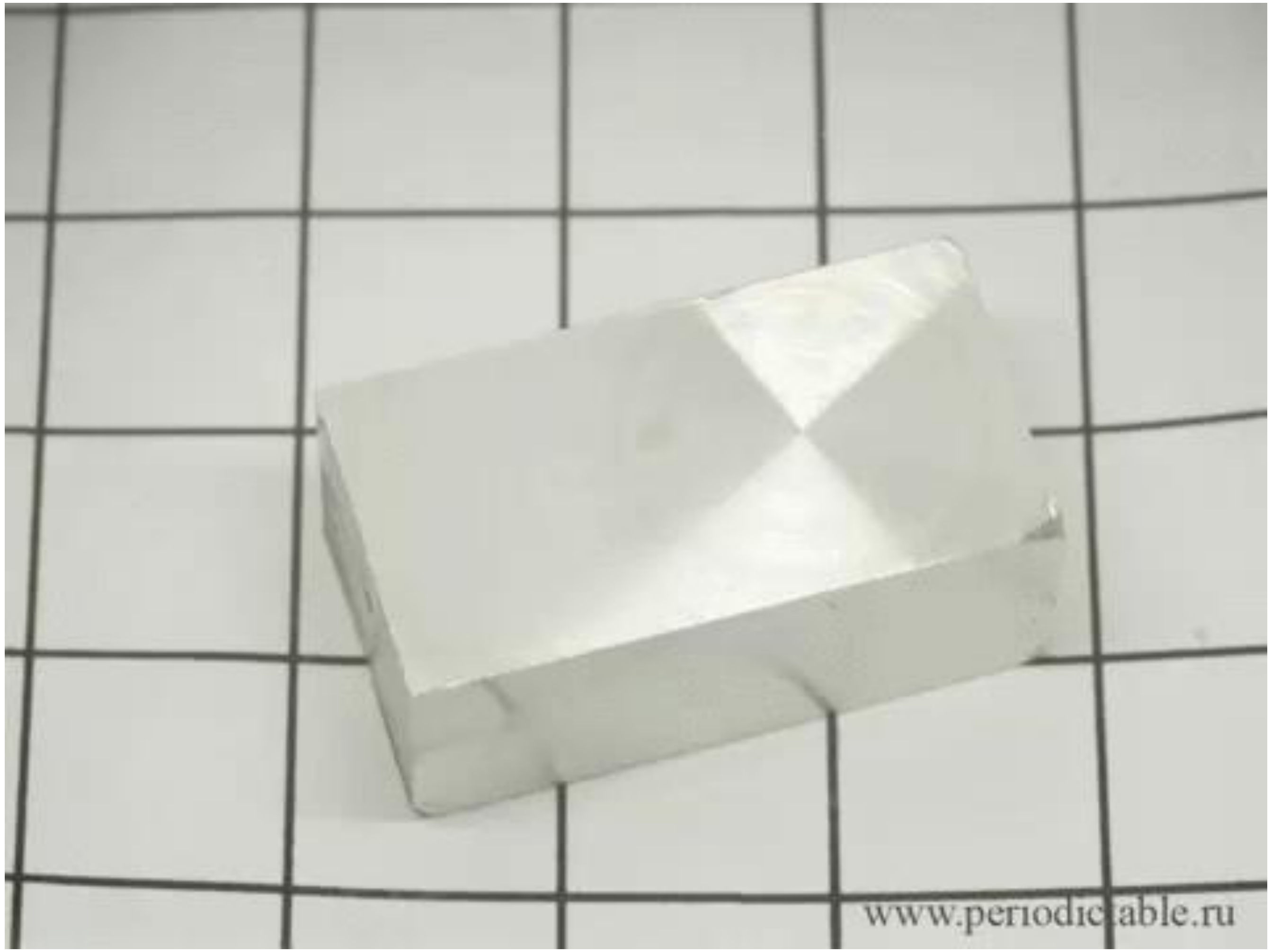
Exp Data: http://res.tagen.tohoku.ac.jp/~waseda/scm/LIQ/periodic_table.html

Application on Tin (Sn)!

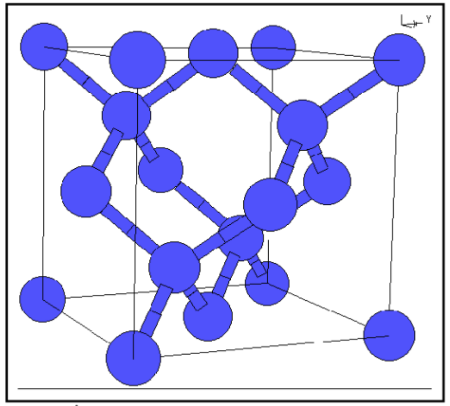
β-Sn to α-Sn phase transition

Beta tin



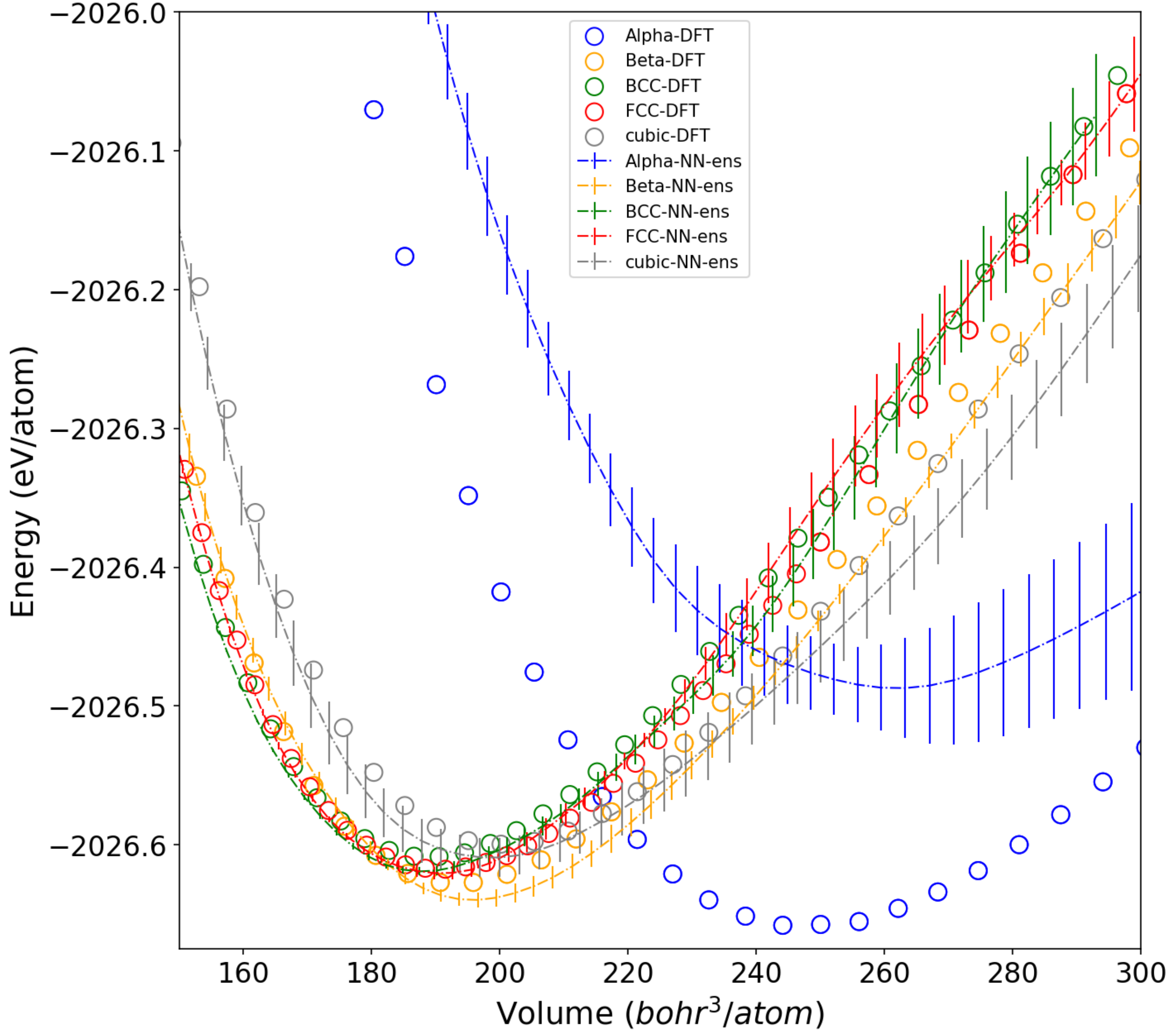


Alpha tin

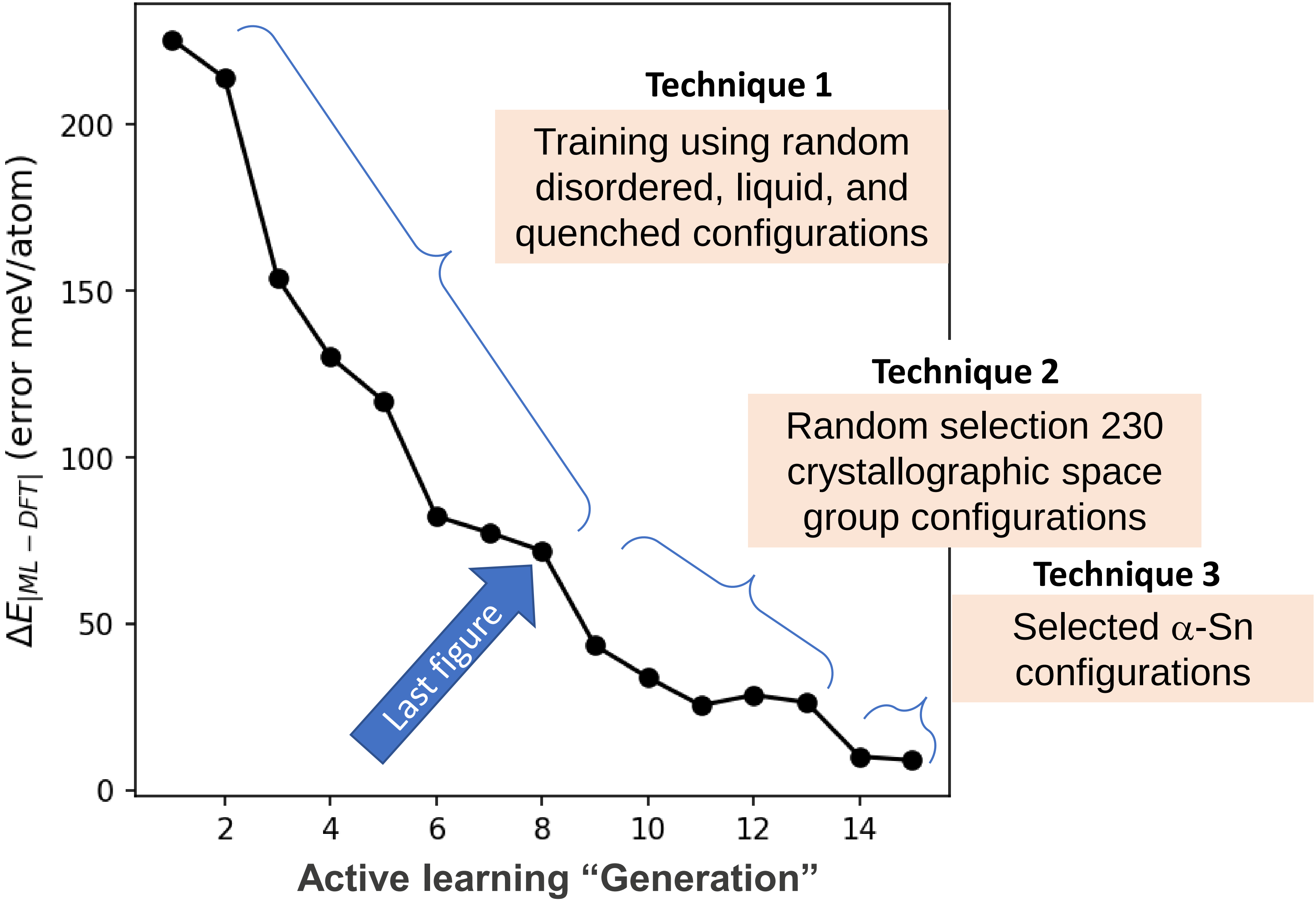


Random disorder MD sampling for Sn

Active learning on Sn w/o any human intervention

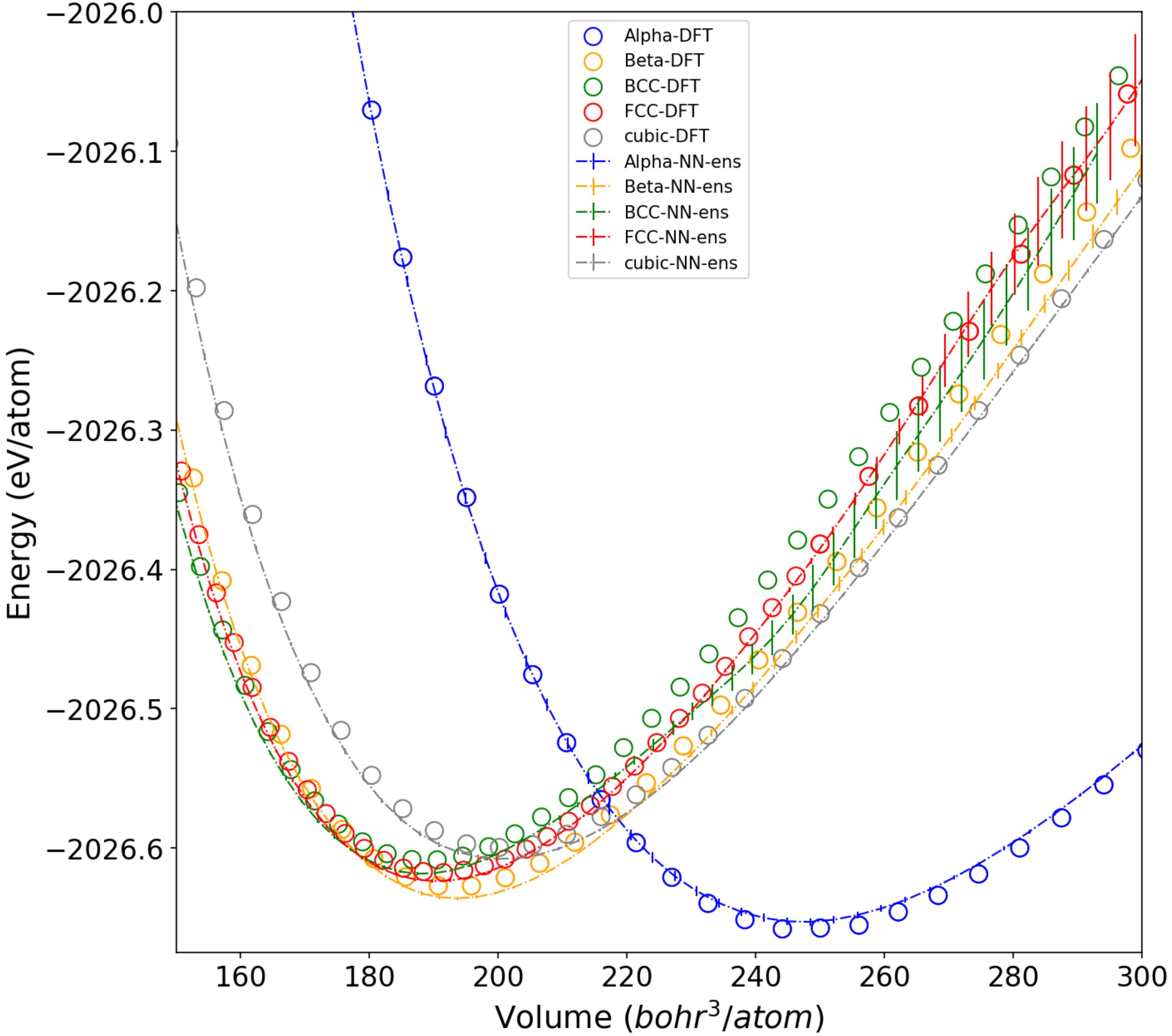


Error on α-Sn with AL progress

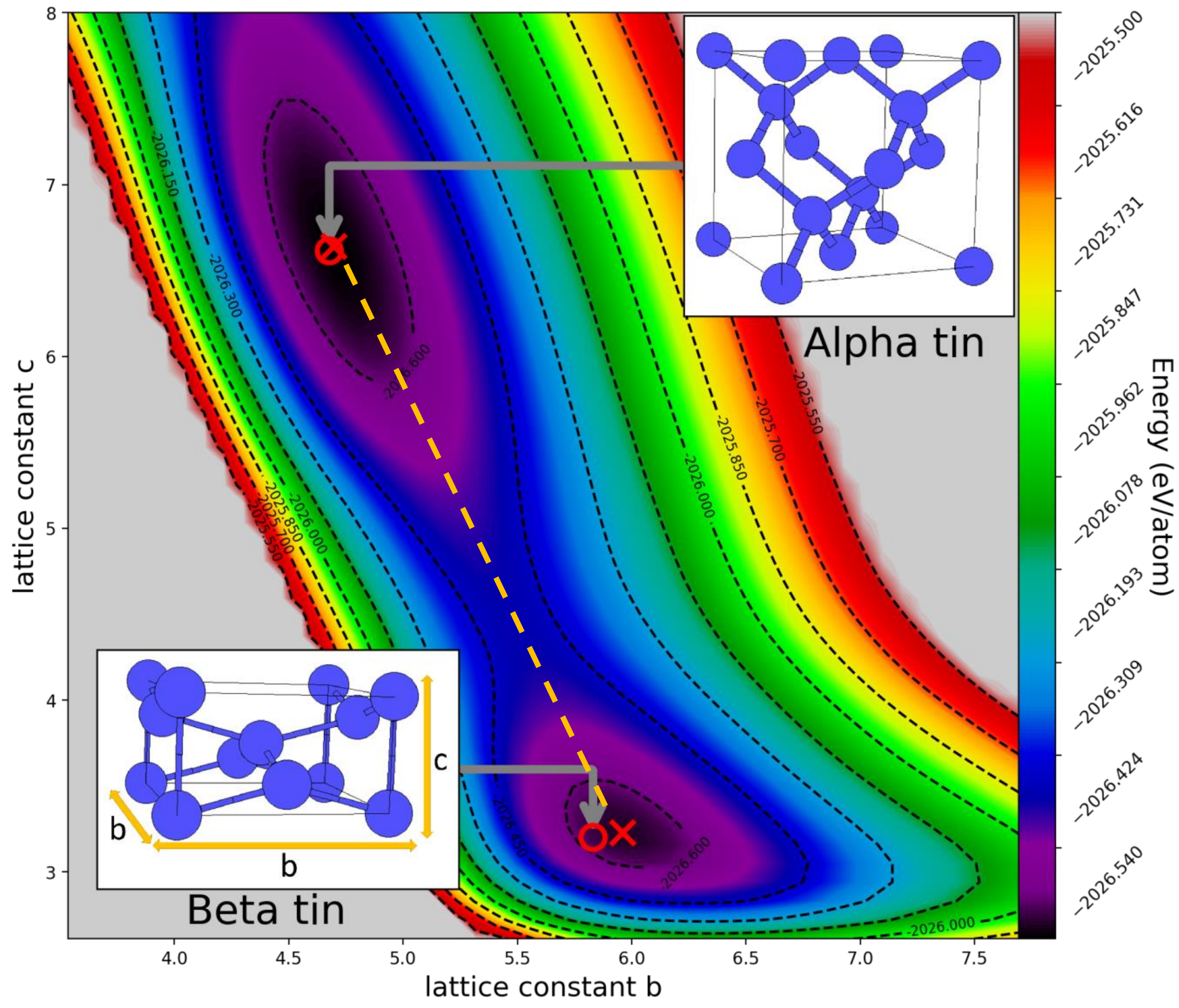


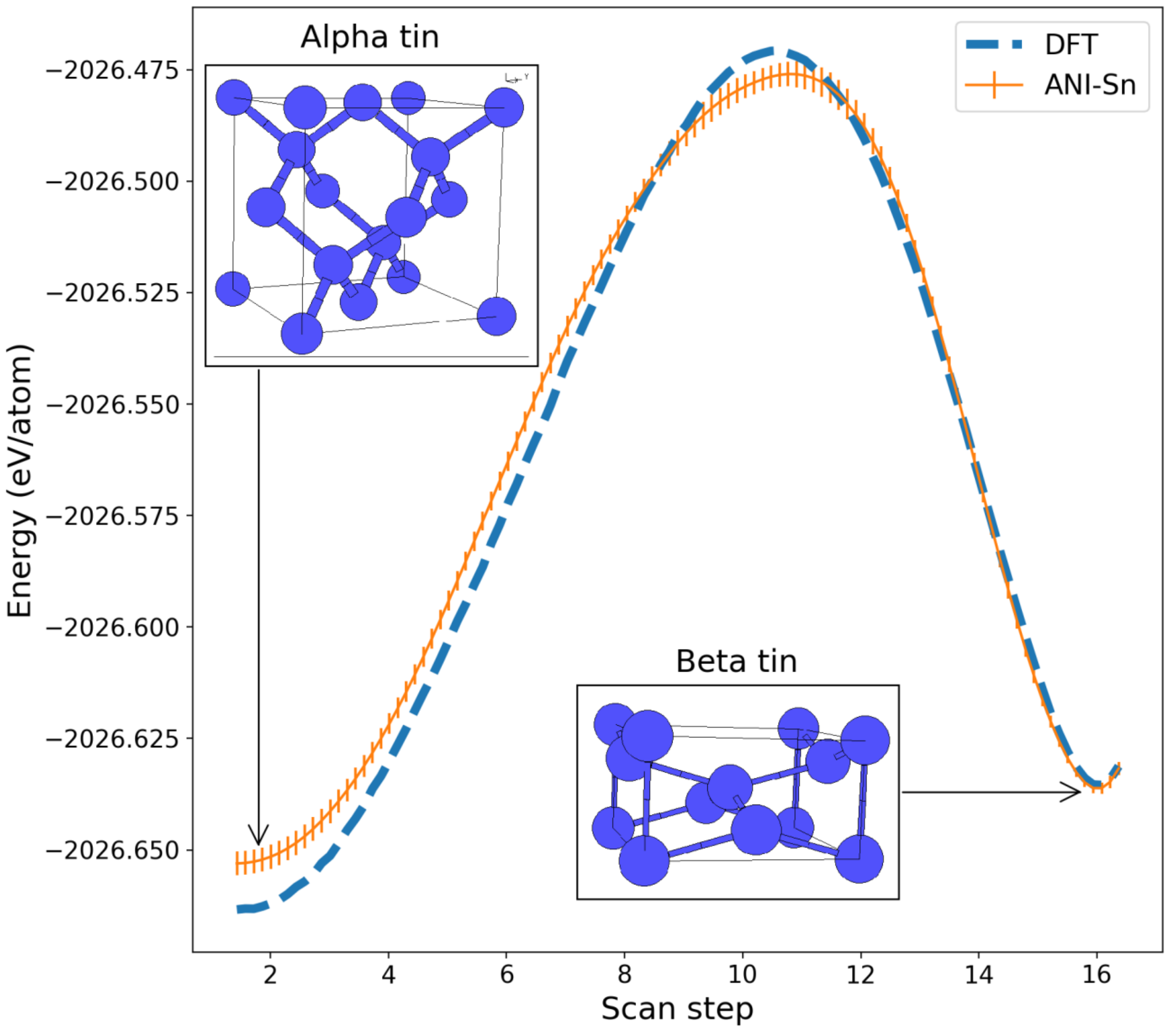
Alpha Sn requires explicit alpha fitting

Including hand picked crystal structures



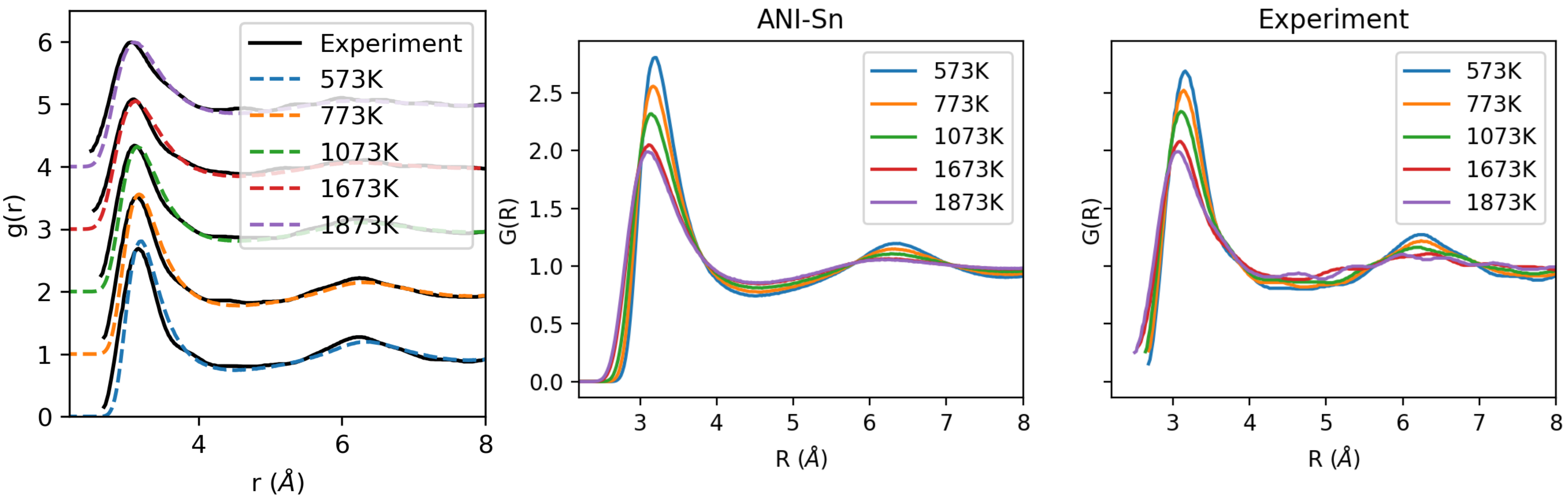
Good agreement with DFT for crystals AND barriers





Liquid Sn RDFs at variable temperatures

- 125ps of NPT for equilibration
- 125ps of NVT at equilibrated density
- 1728 atoms system
- Trained to DFT (PBE)
- Density vs exp ~9% off (on the level of typical DFT error)



Exp data: T. Itami, S. Munejiri, T. Masaki, H. Aoki, Y. Ishii, T. Kamiyama, Y. Senda, F. Shimojo, and K. Hoshino *Phys. Rev. B* 67, 064201 (2003)

How different sampling methods perform

Datasets:

Random = **Random configurations** and **random space groups**
Crystal = **Selected randomly perturbed crystals** using human knowledge

Typical MEAM fitness on Crystals

Energy RMSE (meV/atom)	Force RMSE (eV/A)
44.0	0.97

Testing on Crystals

Training Data	Testing Data	Energy RMSE (meV/atom)	Force RMSE (eV/A)
Crystal	Crystal	3.6	0.04
Random	Crystal	13.7 (17.7/8.3)	0.05 (0.05/0.04)
Random + Crystal	Crystal	4.4	0.03

Testing on Random

Training Data	Testing Data	Energy RMSE (meV/atom)	Force RMSE (eV/A)
Crystal	Random	250.1	1.88
Random	Random	5.9	0.09
Random + Crystal	Random	5.1	0.09

Conclusion and Outlook

Conclusions

- Sampling matters in dynamical studies
- Better data makes a better ML potential
- Active learning methods are required for better data
- Current models may be missing the ability to describe physics in some metals
- Test set results for atomistic ML models can be misleading

Opportunities

- Continue to develop better sampling techniques for metals and molecules
- Discover better uncertainty quantification methods for active learning
- Apply models to gain physical insights
- Recover long range interactions through combined charge prediction and coulomb models

Thank you!

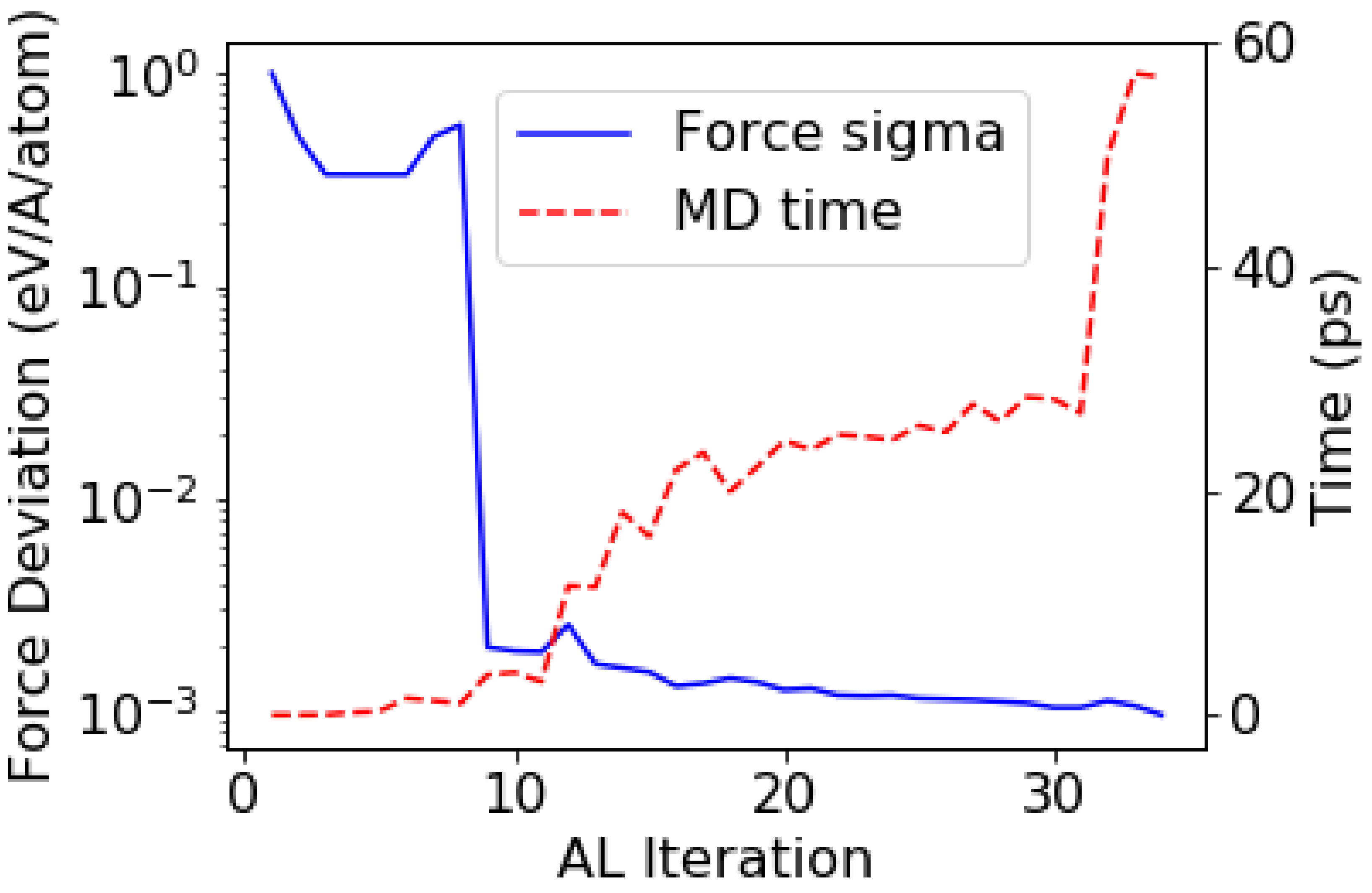
Los Alamos
NATIONAL LABORATORY

Center for Nonlinear Studies
EST. 1980

ASC™

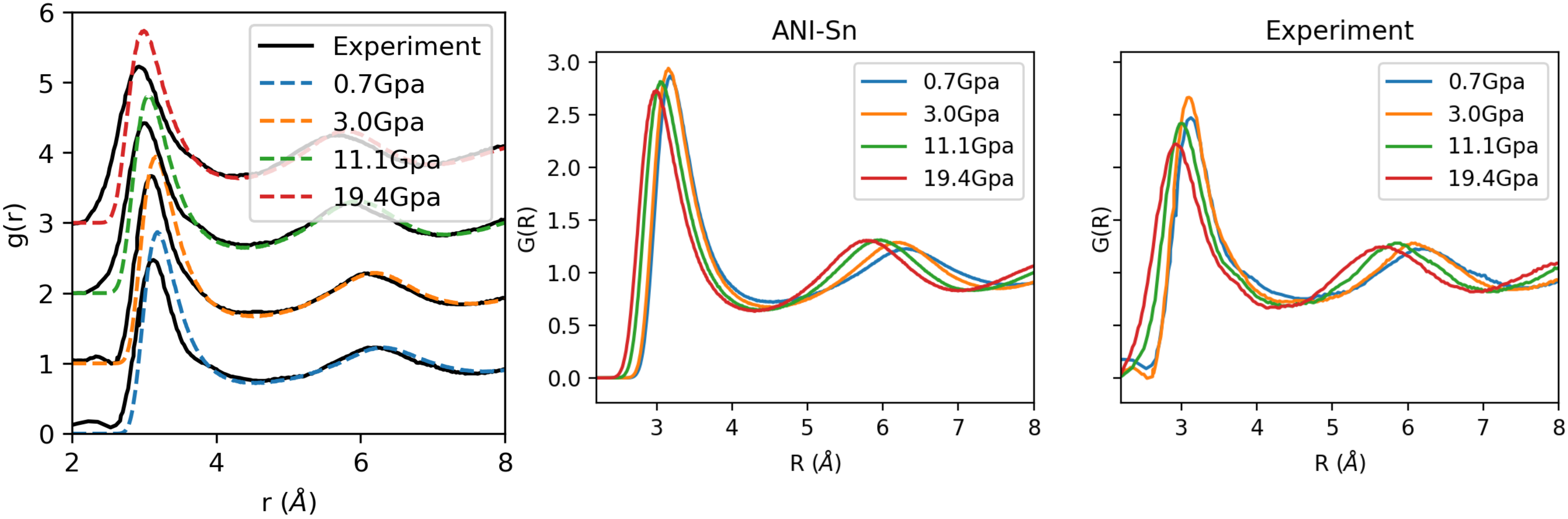
LLNL Sierra
(#2 on TOP500 list)

Thanks for the open science early access allocation!



Force Deviation (eV/A/atom)
10^0
10^{-1}
10^{-2}
10^{-3}
Force sigma
MD time
60
40
20
0
Time (ps)
0
10
20
30
AL Iteration

RDFs at variable pressures

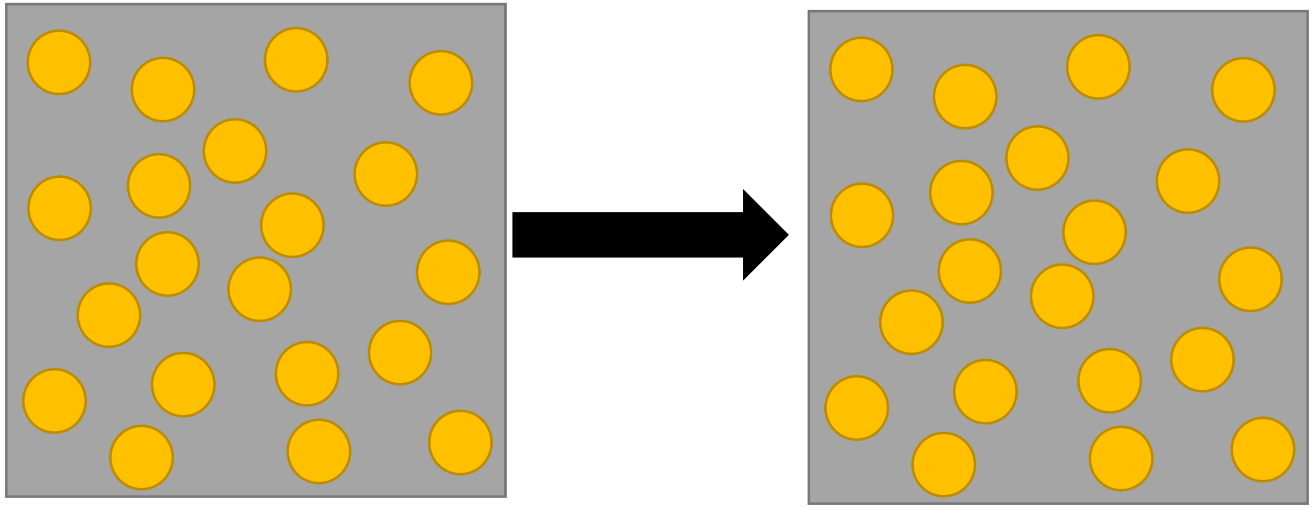


Exp data: T. Narushima, T. Hattori, T. Kinoshita, A. Hinzmann, and K. Tsuji *Phys. Rev. B* 76, 104204 (2007)

Random disorder (no human knowledge) sampling technique

Random atom placement

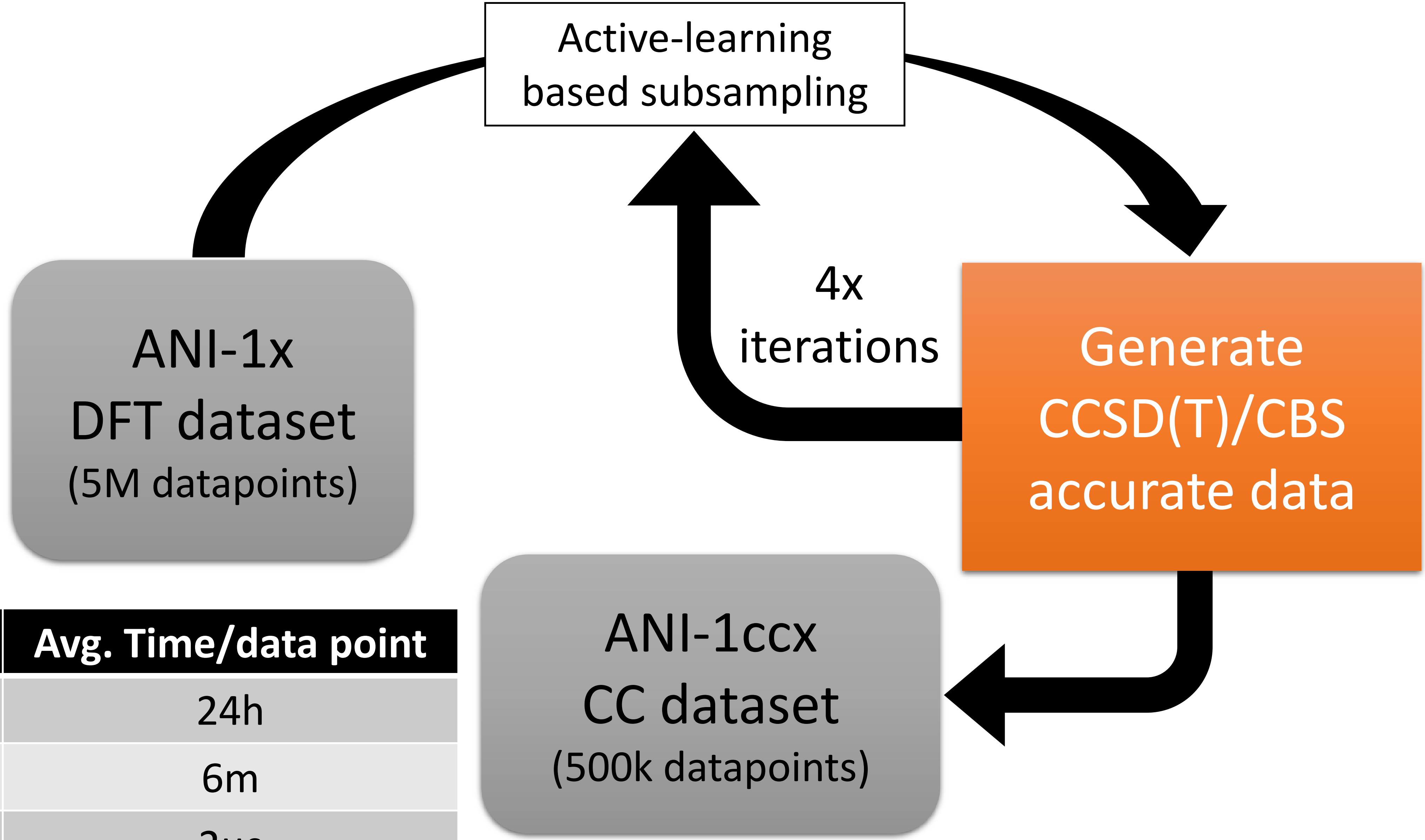
AL MD Sampling with current ML potential



- Minimum atomic distances restrained
- Density kept within a set range
- Random a, b, c lattice constant
- Box size kept minial

- NVT dynamics
- Randomized starting and ending temperature
- Simulation ends with high ensemble disagreement

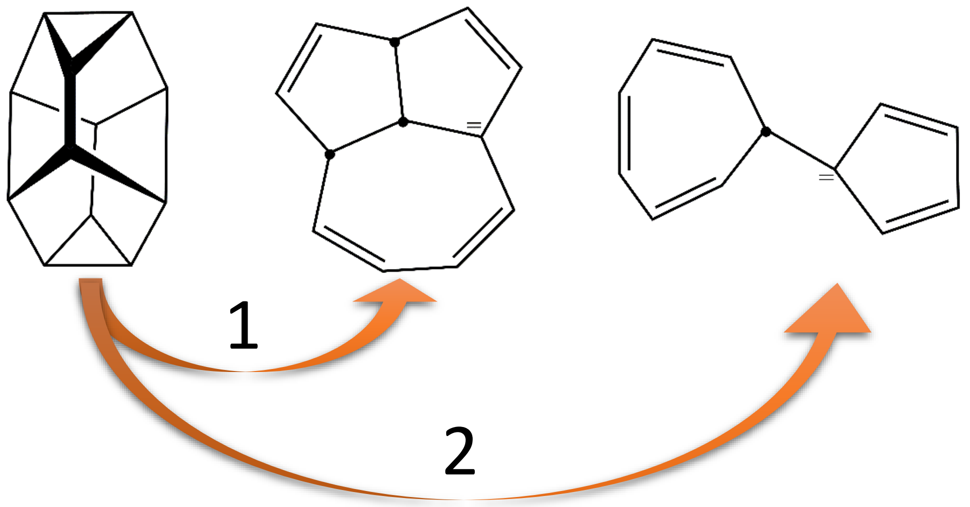
CCSD(T)/CBS accurate data generation

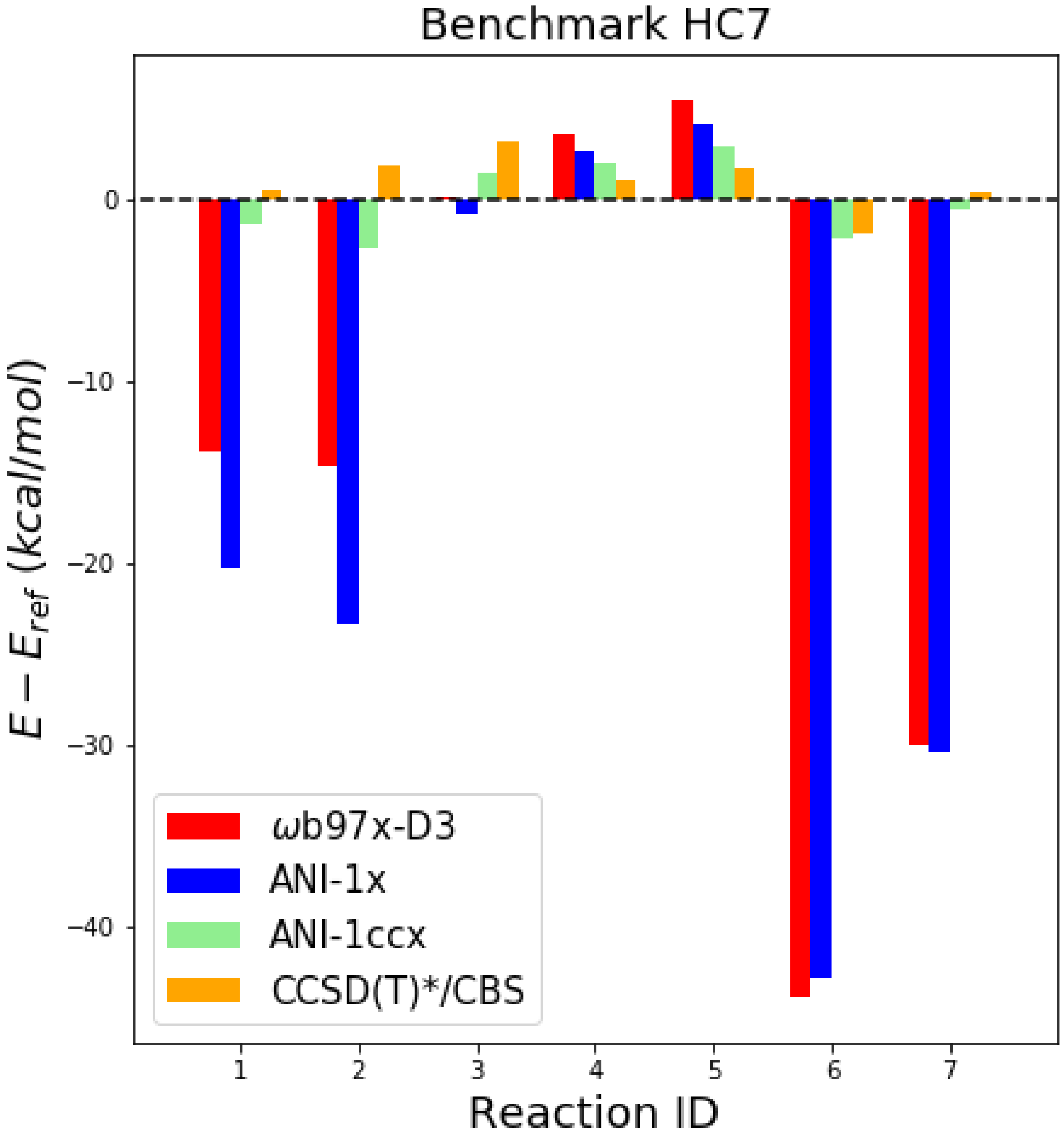


Method	Avg. Time/data point
CCSD(T)	24h
DFT	6m
ANI-1ccx	2µs

Hydrocarbon reaction energy benchmark

Examples





Reference data: Peverati, R.; Zhao, Y.; Truhlar, D. G., *J. Phys. Chem. Lett.* **2011**, *2* (16), 1991–1997.

Organic reaction energy benchmark

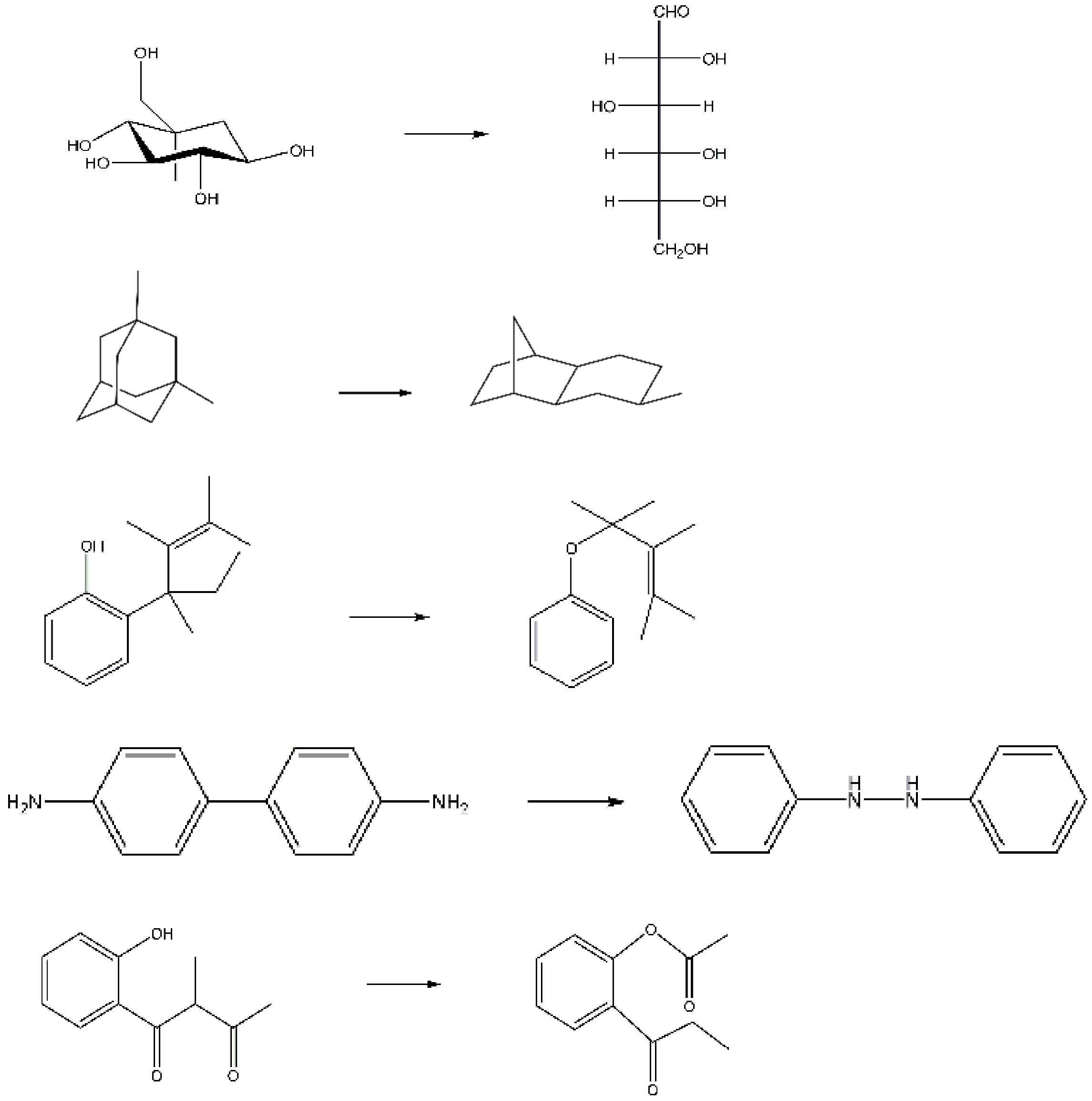
1)

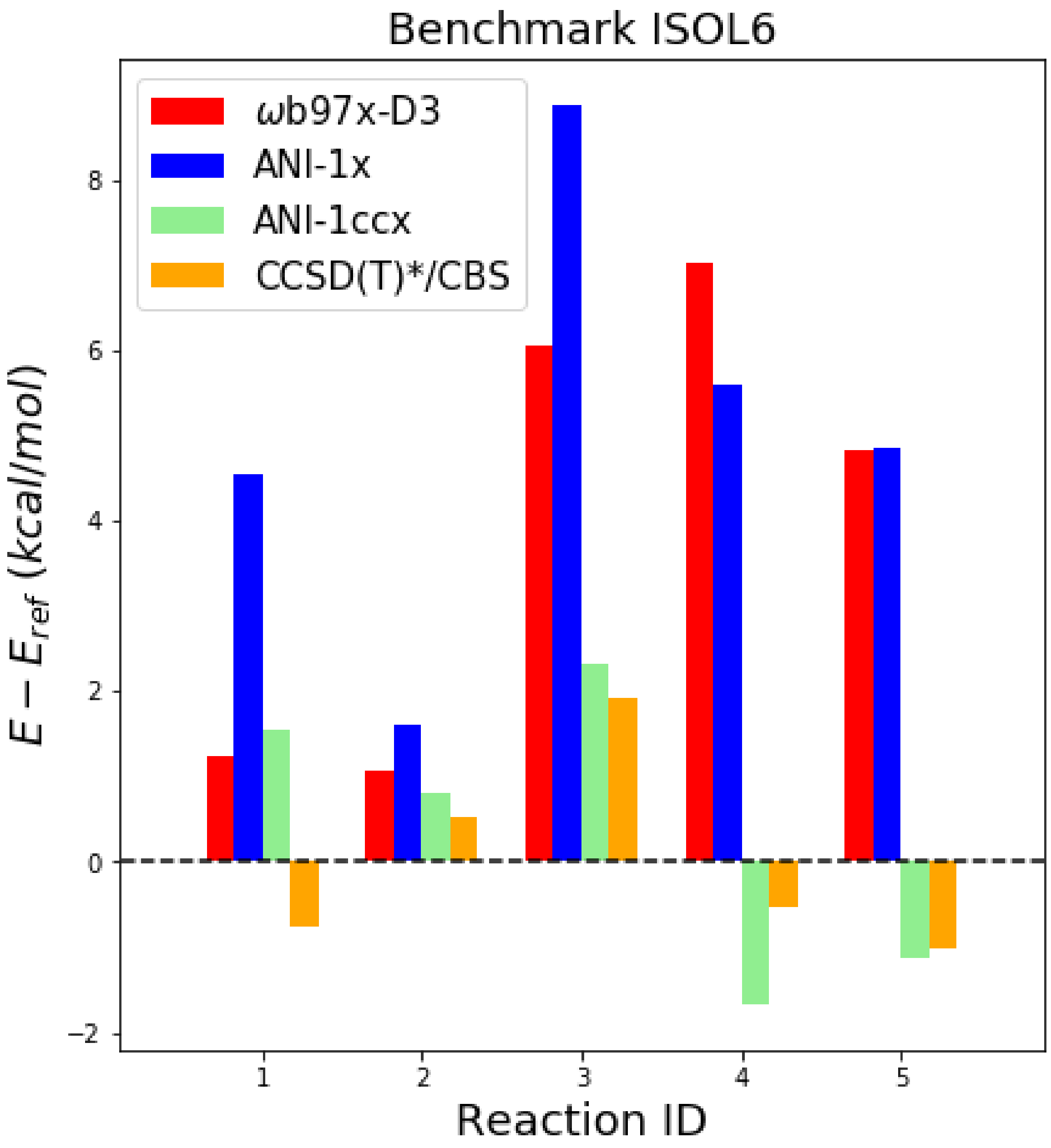
2)

3)

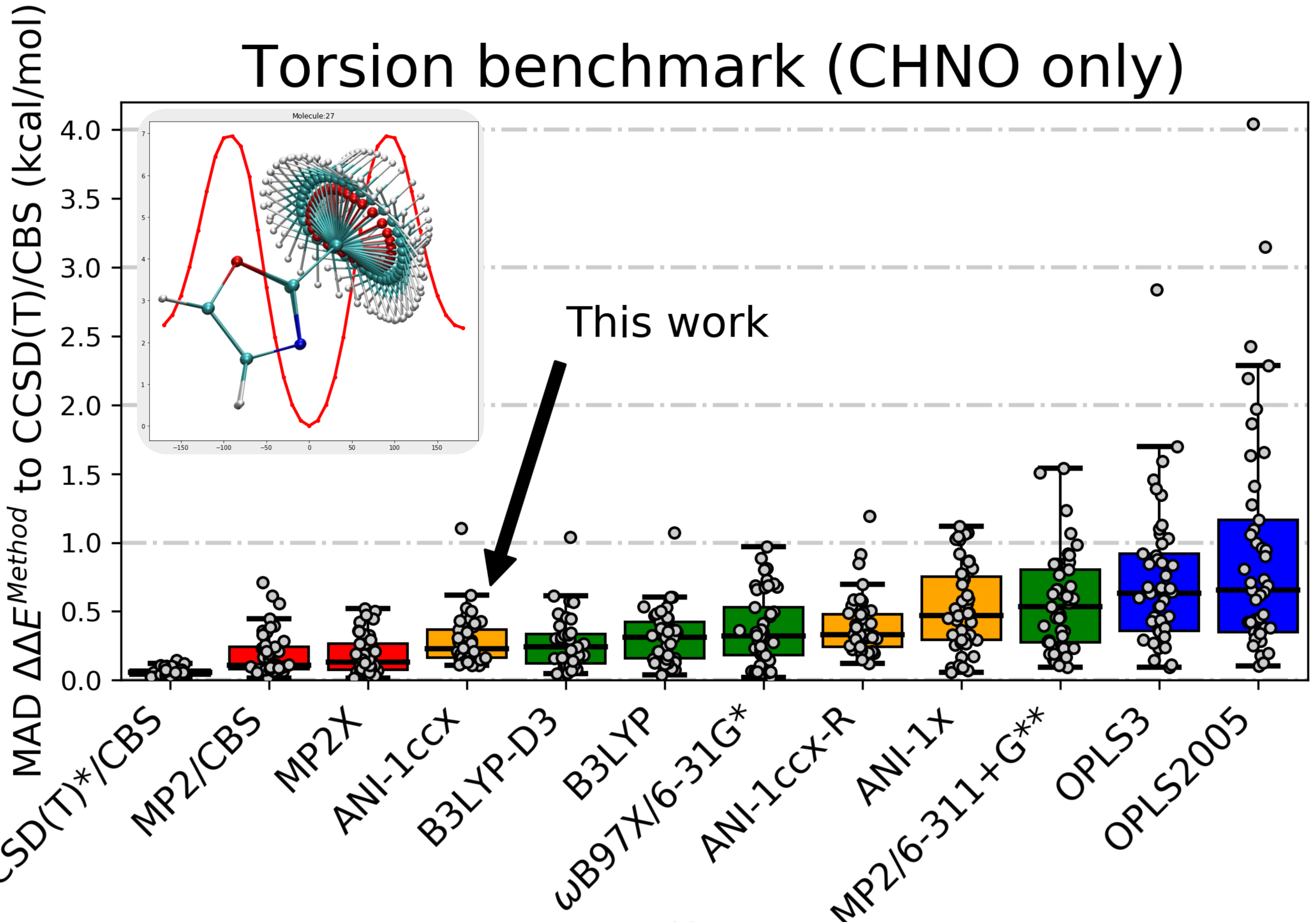
4)

5)



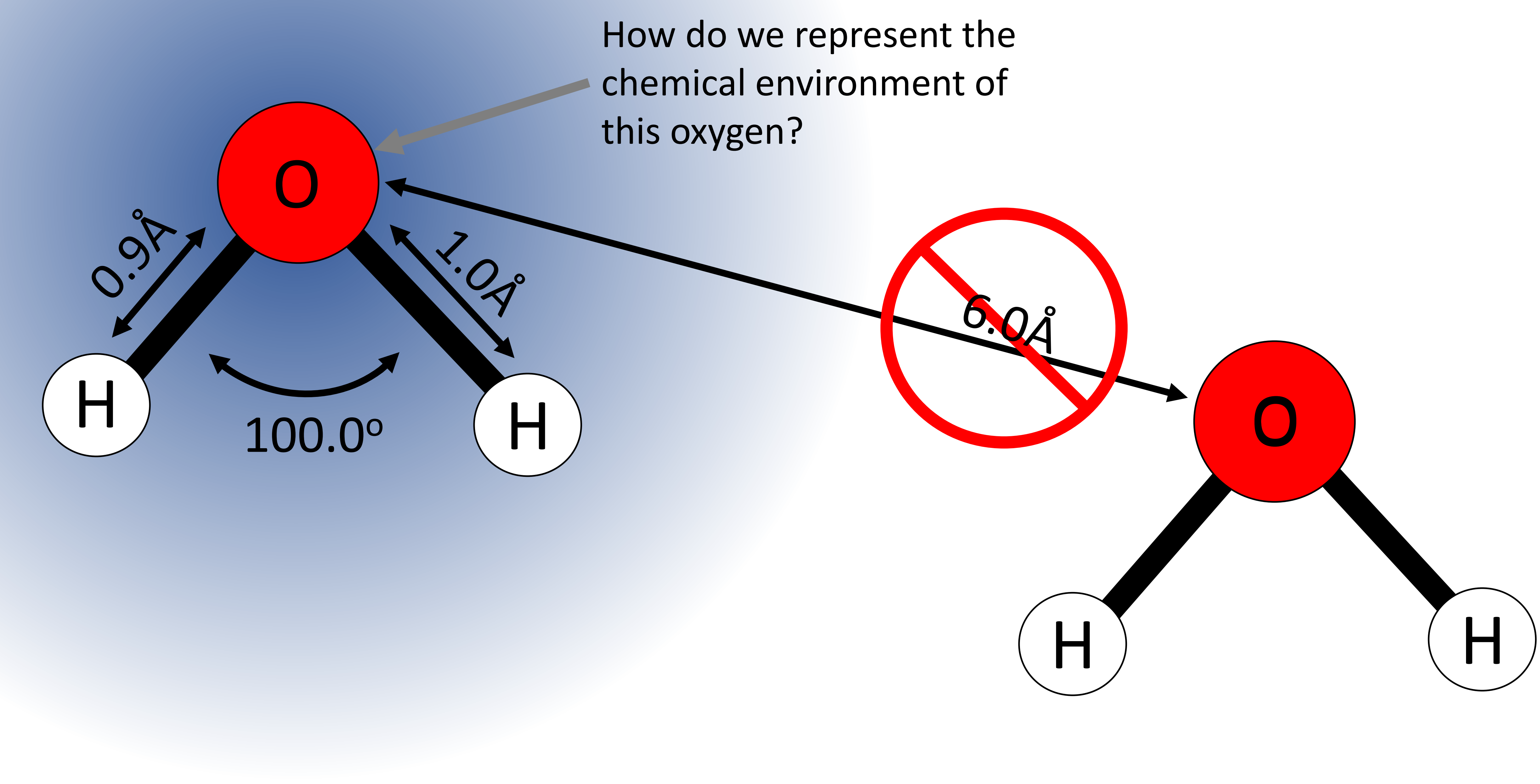


Reference data: Peverati, R.; Zhao, Y.; Truhlar, D. G., *J. Phys. Chem. Lett.* **2011**, *2* (16), 1991–1997.



Benchmark from: Sellers, B. D.; James, N. C.; Gobbi, A., *J. Chem. Inf. Model.* **2017**, *57* (6), 1265–1275.

Atomic environment description



Descriptors for the ANI ML-based potential

Radial symmetry function

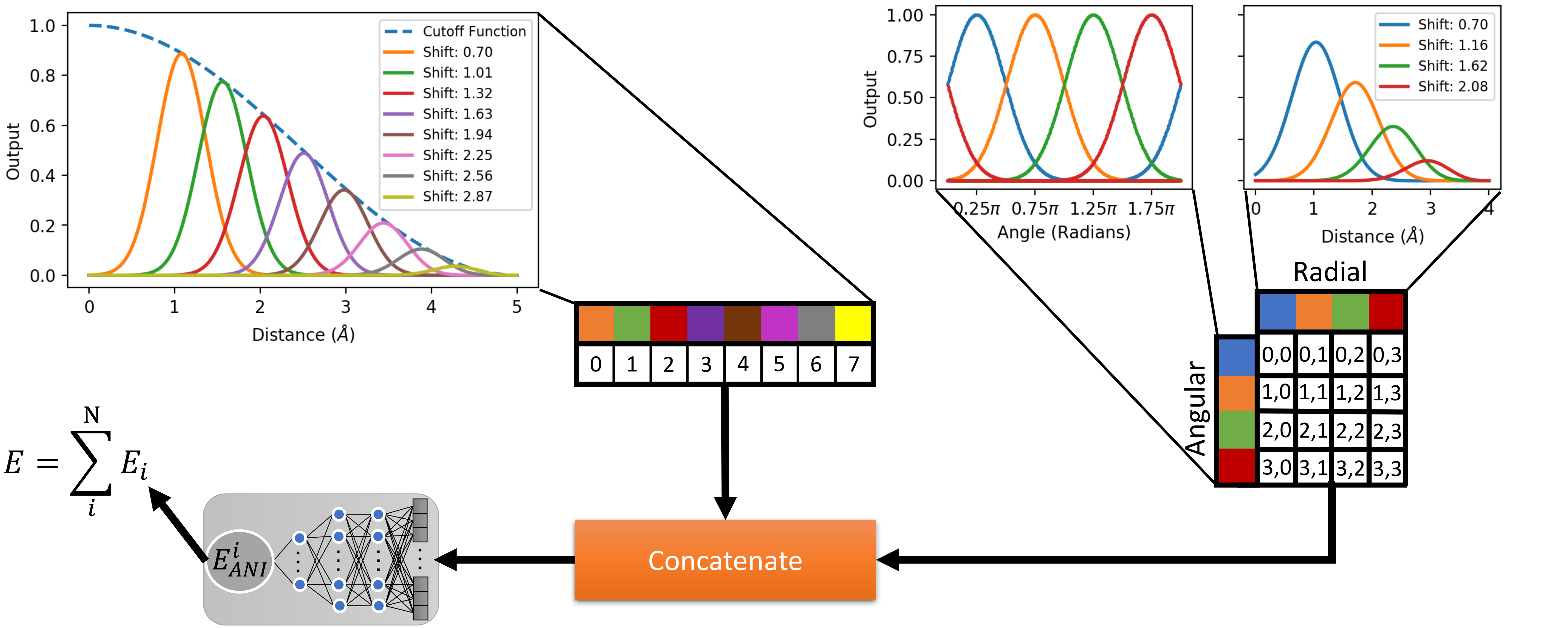
$$G_m^R = \sum_{j\neq i}^{\text{All Atoms}} e^{-\eta(R_{ij}-R_S)^2} f_C(R_{ij})$$

Cutoff function

$$f_C(R_{ij}) = \begin{cases} 0.5 \times \cos\left(\frac{\pi R_{ij}}{R_C}\right) + 0.5 & \text{for } R_{ij} \leq R_C \\ 0.0 & \text{for } R_{ij} > R_C \end{cases}$$

Angular symmetry function

$$G_m^{A\text{mod}} = 2^{1-\zeta} \sum_{j,k\neq i}^{\text{All Atoms}} \left(1 + \cos(\theta_{ijk} - \theta_s)\right)^{\zeta} exp\left[-\eta\left(\frac{R_{ij}^2 + R_{ik}^2}{2} - R_S\right)^2\right] f_C(R_{ij}) f_C(R_{ik})$$



$$E = \sum_{i}^{N} E_i$$